Many-Body Localization in Disordered Quantum Spin Chain and Finite-Temperature Gutzwiller Projection in Two-Dimensional Hubbard Model:

Author: Wei Zhang

Many-Body Localization in Disordered Quantum Spin Chain and Finite-Temperature Gutzwiller Projection in Two-Dimensional Hubbard Model

Wei Zhang

A dissertation

submitted to the Faculty of

the department of Physics

in partial fulfillment

of the requirements

for the degree of

Doctor of Philosophy

Boston College

Morrissey College of Arts and Sciences

Graduate School

August 2019

Many-Body Localization in Disordered Quantum Spin Chain and Finite-Temperature Gutzwiller Projection in Two-Dimensional Hubbard Model

Wei Zhang

Advisor: Chair: Ziqiang Wang, Ph.D., D.Sc.

ABSTRACT

The transition between many-body localized states and the delocalized thermal states is an eigenstate phase transition at finite energy density outside the scope of conventional quantum statistical mechanics. We apply support vector machine (SVM) to study the phase transition between many-body localized and thermal phases in a disordered quantum Ising chain in a transverse external field. The many-body eigenstate energy E is bounded by a bandwidth $W = E_{max} - E_{min}$. The transition takes place on a phase diagram spanned by the energy density $\epsilon = 2(E - E_{min})/W$ and the disorder strength δJ of the spin interaction uniformly distributed within $[-\delta J, \delta J]$, formally parallel to the mobility edge in Anderson localization. In our study we use the labeled probability density of eigenstate wavefunctions belonging to the deeply localized and thermal regimes at two different energy densities (ϵ's) as the training set, *i.e.*, providing labeled data at four corners of the phase diagram. Then we employ the trained SVM to predict the whole phase diagram. The obtained phase boundary qualitatively agrees with previous work using entanglement entropy to characterize these two phases. We further analyze the decision function of the SVM to interpret its physical meaning and find that it is analogous to the inverse participation ratio in configuration space. Our findings demonstrate the ability of the SVM to capture potential quantities

that may characterize the many-body localization phase transition.

To further investigate the properties of the transition, we study the behavior of the entanglement entropy of a subsystem of size L_A in a system of size $L > L_A$ near the critical regime of the many-body localization transition. The many-body eigenstates are obtained by exact diagonalization of a disordered quantum spin chain under twisted boundary conditions to reduce the finite-size effect. We present a scaling theory based on the assumption that the transition is continuous and use the subsystem size L_A/ξ as the scaling variable, where ξ is the correlation length. We show that this scaling theory provides an effective description of the critical behavior and that the entanglement entropy follows the thermal volume law at the transition point. We extract the critical exponent governing the divergence of ξ upon approaching the transition point. We again study the participation entropy in the spin-basis of the domain wall excitations and show that the transition point and the critical exponent agree with those obtained from finite size scaling of the entanglement entropy. Our findings suggest that the many-body localization transition in this model is continuous and describable as a localization transition in the many-body configuration space.

Besides the many-body localization transition driven by disorder, We also study the Coulomb repulsion and temperature driving phase transitions. We apply a finite-temperature Gutzwiller projection to two-dimensional Hubbard model by constructing a "Gutzwiller-type" density matrix operator to approximate the real interacting density matrix, which provides the upper bound of free energy of the system. We firstly investigate half filled Hubbard model without magnetism and obtain the phase diagram. The transition line is of first order at finite temperature, ending at 2 second order points, which shares qualitative agreement with dynamic mean field results. We derive the analytic form of the free energy and therefor the equation of states, which benefits the understanding of the different

phases. We later extend our approach to take anti-ferromagnetic order into account. We determine the Neel temperature and explore its interesting behavior when varying the Coulomb repulsion.

Dedicated to my family

ACKNOWLEDGEMENTS

I would like to thank everyone who has been part of my life during these years at Boston College. They make this time at Boston College precious and enjoyable. First of all, I would like to thank my advisor Professor Ziqiang Wang. He taught me how to do research and to think about and solve problems. His deep thought and curiosity to physics has been constantly inspiring me to explore challenging problems. When I had questions, he was always there for me with his knowledge, creativity, guidance and patience. He also gave me freedom to choose project according to my own interest, and provided me great support and help to realize my ideas. I learned a lot from him, which will keep benefiting my future life. I am very happy and grateful to be his student. I would also like to thank all professors who taught me in class where I truly enjoyed the beauty of physics, and whom I worked for as a teaching assistant: Professor Broido, Professor Kempa, Professor Ran, Professor Bakshi, Professor Kalman, Professor Uritam, Professor Bedell, Professor Engelbrecht, Professor Herczynski, Professor Graf, Professor Di Bartolo, Professor He. And many thanks to Jane, Nancy, Sile and Scott for helping in all aspects of my life in Boston College.

During the years, I have made great friends in Boston College. Thanks to Yi Zhang, Kun Jiang, Bing Ye, Matthew Gochan, Matthew Heine, Luciano Silvestri, Thomas Mion, Joshua Heath and lots of other people, but too many to name. I have many joyful memories with them.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

CHAPTER I

Introduction

The study of phase transitions is one of the most important aspects in physics. From every-day examples like boiling water and steam to more exotic ones like superconductivity and superfluidity, the nature of phase transitions is so rich, making that how to identify, classify and understand them subtle questions.

In the modern classification, phase transitions are divided into two broad categories, namely first-order phase transition and second-order phase transition. For the former, a familiar example is the boiling of water, where the water does not instantly turn into vapor, but forms a turbulent mixture of liquid water and vapor bubbles, which illustrates a characteristic property of the first-order phase transition: phase coexistence across the transition. Another more sophisticated example involves the U-T phase diagram (U is the interaction strength, T is the temperature) of the frustrated Hubbard model in the limit of large lattice coordination, known as Mott transition which will be detailed in later chapters: there exists a first-order phase-transition line at nonzero temperature[2]. The first-order phase transition is usually characterized by a discontinuity in the first derivative of the free energy with respect to some thermodynamic variables.

In contrast, the second-order phase transition does not involve phase coexistence or discontinuity in the first derivative of the free energy, it's continuous in

the first derivative, but non-analytical in the second derivative of free energy. The study of second-order phase transition has been one of the most fascinating area of modern physics since last century, especially enlightened by Landau-Ginzburg theory and concepts of spontaneous symmetry breaking. One famous example is the Curie point of a ferromagnet. The Curie point is the critical point T_c, when $T < T_c$, the magnetic moment of the system $m > 0$, while when $T > T_c$, m vanishes. This transition is driven by thermal fluctuations. The correlation length ξ, namely the length scale determining the exponential decay of the two-point correlation function, diverges when $T \to T^+$ as $\xi \sim (T - T_c)^{-\nu}$, where ν is the correlation length exponent which is universal. In addition to this example, the Mott transition mentioned above, while it has a first-order phase-transition line in the U-T phase diagram, the line ends in a second-order (quantum) critical point at zero temperature.

A subclass of phase transitions which occur only at zero temperature, is quantum phase transition (QPT). Differing from the classical phase transitions, QPTs are purely driven by quantum fluctuations instead of thermal fluctuations. They can also be classified to be first-order or second-order. The first-order QPTs have discontinuities in the first derivative of the ground state energy density. The second-order ones, being continuous in the first derivative, have non-analytical second derivatives in most cases. They are usually signaled by order parameters vanishing continuously at the quantum critical point, namely, at some critical values of parameters characterizing the Hamiltonian, such as interaction strength, magnetic field, pressure, and so on.

Great progress has been made in understanding QPTs in recent decades. Instead of taking a thorough look at all of them, we briefly discuss two critical exponents which are more relevant to our later results. Denote the parameter characterizing the Hamiltonian as g, when $g \to g_c$, the energy gap Δ between

the ground state and the first excited state vanishes, the system become gapless at the critical point g_c. In the vicinity of the transition, $\Delta \sim |g - g_c|^{z\nu}$, where z, ν are critical exponents. Besides, in analogy to the correlation length for the second-order classical phase transition, here ξ is the correlation length defined in a similar manner, by $G(r) = (\langle O(0,t)O(r,t)\rangle - \langle O(0,t)\rangle\langle O(r,t)\rangle) \sim \frac{e^{-r/\xi}}{r^{d-2+\eta}}$,which is the equal-time correlation function, and η is the Fisher exponent and d is the dimensionality of the system. ξ diverges as $\xi \sim |g - g_c|^{-\nu}$ when $g \to g_c$. As a result, $G(r)$ decays as a power law instead of exponentially at transition.

People may argue that as zero temperature is impossible to be reached in experiments, the study of QPTs does not shed light to understanding real physical systems, however, the knowledge about QPTs are critical to understand the behavior of systems in the quantum critical region at finite temperature.

Another important class of phase transitions is the eigenstate phase transitions, of which many-body localization (MBL) phase transition is an illustrating example. We will focus on this topic in later chapters. Here for completeness we mention that in contrast to ordinary quantum phase transitions that occur in ground states, the MBL phase transition is at finite energy densities. It falls outside of the frame of conventional quantum statistical mechanics which averages over many eigenstates, due to the breakdown of ergodicity in the MBL phase. As a result, its nature is not yet clear now.

This work is composed of three parts. We (1) study the MBL transition using machine learning approach, the machine learning results implies that MBL is a genuine localization in spin configuration space, which supports an argument still under debate. We (2) explore the critical behavior of entanglement entropy and participation entropy near many-body localization transition by developing a scaling theory and perform finite size scaling based on that theory, the results suggests the continuity of MBL phase transition, helping understand the funda-

mentals of this unconventional eigenstate phase transition. We then (3) presents a study of Mott transition at finite temperature using Gutzwiller approximation. We generalize the Gutzwiller projection from ground states to finite temperature and determine the U-T phase diagram, and equations of states, which helps to understand the Mott phase and phase transition at finite temperature. The results are compared with DMFT results.

Below I briefly review some key concepts of many-body localization. Then I briefly introduce the foundation of Gutzwiller approximation, specifically, Gutzwiller approximation of Hubbard model. In both part, I focus on basics, I refer the readers to materials[3–6] for MBL and materials[7–9] for Gutzwiller approximation, which present more comprehensive and extensive study and details.

1.1 Many-Body Localization

1.1.1 Anderson Localiztion

In 1958, Anderson published the well-known paper where he discussed the electrons' behavior in crystal with impurities. In contrast to the behavior in clean system, the electron in such disordered system is trapped due to the external random potential, its motion is no longer diffusive, so even after infinite long time, its wavefunction will not be extensive, instead, it will take a characteristic exponential shape in space, as shown in Fig. 1.1. That is to say, the wavefunction that takes the asymptotic long distance form $\Psi_n(\mathbf{r}) \sim \exp(-\frac{|\mathbf{r}-\mathbf{R}_n|}{\xi})$ is localized, where ξ here is the localization length, n denotes the n^{th} state localized near position $\mathbf{R}_n$. In 1D and 2D, all eigenstates become localized at infinitely small disorder. But in 3D systems, the transition between the localized and extended states can happen. Below the critical disorder strength, there exists an extensive regime in the middle of the band. This regime is separated with the rest of band

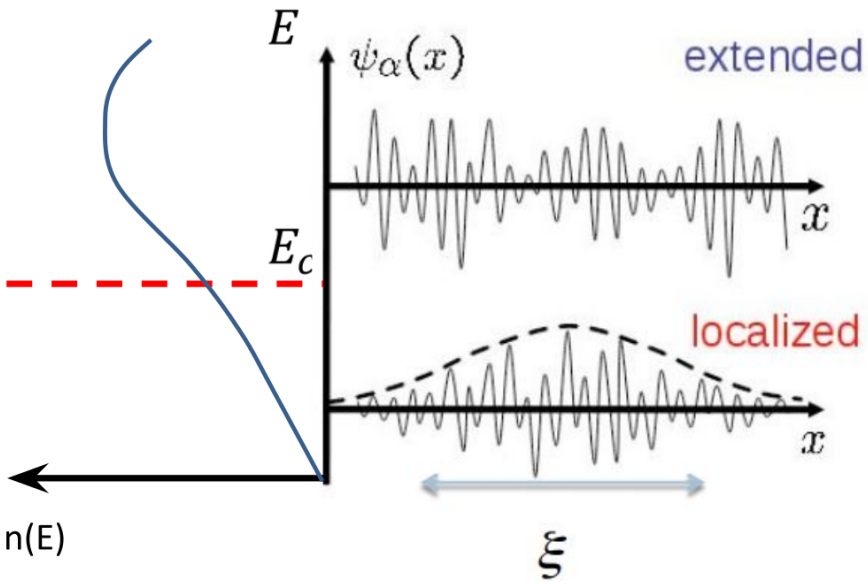

Figure 1.1: A schematic plot of Anderson localization. When the disorder strength is smaller than the critical value, there exists mobility edges in the band separating localized eigenstates from extensive ones.

by the mobility edge E_c. There are two mobility edges in the band (though only one is shown in Fig. 1.1 for simplicity). When the disorder strength increases, the mobility edge moves toward the band center, until at the critical disorder strength, the extensive regime disappears, all eigenstates become localized.

A simple example of Hamiltonian displaying Anderson localization transition can be a tight-binding model of a single quantum particle hopping on an infinite lattice:

$$\hat{H} = -t \sum_{\langle i,j \rangle} c_i^\dagger c_j + \sum_i U_i c_i^\dagger c_i \tag{1.1}$$

where $\langle i, j \rangle$ denotes pairs of nearest neighbor lattice site and U_i is a static random onsite potential. eq. 1.1 can be transformed to $\hat{H} = \sum_n E_n c_n^\dagger c_n$, where $c_n^\dagger$ creates a particle in the single-particle eigenstate $|n\rangle$. If a particle is initialized to be in a localized $|n\rangle$, it cannot traverse the whole lattice even after infinitely long time.

There are several quantities that are commonly used to characterize the Anderson localized state, including the conductance, the inverse participation ratio

(IPR) the typical density of states and so on. Here we briefly mention IPR for later use.

IPR measures the inverse spatial coverage of a single-particle eigenstate, defined as:

$$I_q(E_n) = \sum_{\mathbf{r}} |\Psi_n(\mathbf{r})|^{2q} \tag{1.2}$$

The size dependence of IPR enables us to distinguish between the metallic and localized regime. In the localized regime, we expect that only a few lattice sites are occupied with $|\Psi_n| \sim 1$. Therefore $I_q \sim 1$ does not depend on the size of the system. In the metallic regime, electron occupies all sites of the lattice, $|\Psi_n(\mathbf{r})|^2 \sim L^{-d}$ where d is the dimension of the system, and $I_q(E_n) \sim L^{-d(q-1)}$. At the critical point, where the wave function is multifractal[10], and $I_q \sim L^{-d_q(q-1)}$ where d_q is the multifractal dimension depending on the value of q. In Chapter IV this single-particle IPR will be generalized to many-body case and be used to study the many-body localization.

More recently, there are also attempts to study the Anderson localization using a single particle quantity called typical density of states, which is defined as the geometric average of the local density of states over the disorder configurations [11, 12]. This method has also been applied to realistic materials to study the Anderson localization for functional materials[13–15].

1.1.2 Thermalization and Eigenstate Thermalization Hypothesis

Basko[6] presented that the localization keeps to remain when weak interactions exist, this is the many-body localization (MBL). In contrast to the delocalized system whose long-time behaviour obeys equilibrium thermodynamics, in MBL phase, the system fails to thermalize and there is a breakdown of ergodicity. Therefore the phase transition between MBL and the delocalized phase is a dynamical phase transition that cannot be described by the conventional quantum

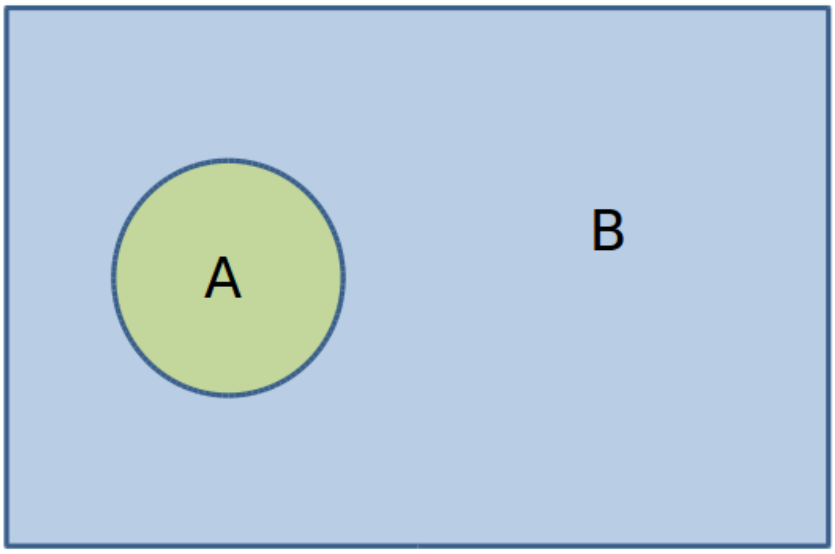

Figure 1.2: A closed quantum system partitioned into a subsystem A and everything else B, while B is much larger than A.

statistical mechanics. Before moving on to more details, we firstly clarify the idea of thermalization in a closed quantum system and the concept of eigenstate thermalization hypothesis (ETH) which is true in the delocalized systems, but not in the MBL phase.

First consider a system satisfying thermodynamic limit, shown in Fig. 1.2, it's partitioned into a subsystem of A, and everything else B. Any choice of subsystem A is acceptable, as long as the degrees of freedom within A as a fraction of the full system's degrees of freedom goes to zero. The whole system in its initial (pure or mixed) state is described by a density matrix $\rho(0)$. After time t, it evolves to $\rho(t) = e^{-i\hat{H}t/\hbar}\rho(0)e^{i\hat{H}t/\hbar}$. The same system at equilibrium at temperature T has Boltzmann density matrix $\rho^{eq}(T) = exp(-\hat{H}/kT)/Z$, where $Z = Tr(e^{-\hat{H}/kT})$ is the partition function, and the temperature T is determined by $Tr(\hat{H}\rho(0)) = Tr(\hat{H}\rho^{eq}(T))$, as the whole system is closed, no energy exchange with the external environment.

The reduced density matrix of the subsystem A at time t is $\rho_A(t) = Tr_B(\rho(t))$.

Also as a subsystem of the whole, the density matrix of A at equilibrium temperature T will have density matrix $\rho_A^{eq}(T) = Tr_B(\rho^{eq}(T))$, where T is defined above.

Thermalization of a closed quantum system means that after infinitely long time evolution, the state of the subsystem should be the same to that if it's coupled to a infinitely large heat bath at temperature T, because the rest of the whole system, B, acts as that heat bath. That is to say, all degrees of freedom in any choice of subsystem A are interacting with the rest of the system, we cannot find any subsystem which isolated from the rest. For any initial state of the system $\rho(0)$ this leads to:

$$lim_{t\to\infty}\rho_A(t) = \rho_A^{eq}(T) \tag{1.3}$$

Note that thermodynamic is assumed for eq. 1.3.

If a system can thermalize, it suggests naturally that all its many-body eigenstates of $\hat{H}$ are thermal, which is the statement of eigenstate thermalization hypothesis (ETH). Consider the system is in an eigenstate $|n\rangle$ satisfying $\hat{H}|n\rangle = E_n|n\rangle$, where E_n is the eigenstate energy corresponding to equilibrium temperature T_n such that $E_n = Tr(\hat{H}\rho^{eq}(T_n))$. As $\rho(0) = \rho^{(n)}(0) = |n\rangle\langle n|$, its time evolution is trivial, $\rho^{(n)}(t) = e^{-i\hat{H}t/\hbar}|n\rangle\langle n|e^{i\hat{H}t/\hbar} = \rho^{(n)}(0), \forall t$, thus leading to $\rho_A^{(n)}(t) = Tr_B\rho^{(n)}(t) = Tr_B\rho^{(n)}(0) = \rho_A^{(n)}(0), \forall t$. Because the system can thermalize, meaning that we have $lim_{t\to\infty}\rho_A^{(n)}(t) = \rho_A^{eq}(T)$, ETH asserts that:

$$\rho_A^{(n)}(0) = \rho_A^{eq}(T) \tag{1.4}$$

Although it's impossible in experiment to initialize a system to be in its eigenstate, ETH is an important tool for studying the MBL phase and MBL-ETH phase transition, as ETH is true in delocalized systems but false in MBL phase. These aspects will be detailed in ChapterII.

A significant consequence of ETH is the volume law of the subsystem entanglement entropy. The entanglement entropy of subsystem A in eigenstate $|n\rangle$ is defined as:

$$S_E^A(n) = -k_B Tr_A(\rho_A^{(n)} log \rho_A^{(n)}) \tag{1.5}$$

Plugging eq. 1.4 into eq. 1.5, we have $S_E^A(n) = S^A(T_n)$, namely, the entanglment entropy equals to equilibrium thermal entropy of subsystem A at finite temperature T_n. Because thermal entropy is an extensive property proportional to system size, it implies that the entanglement entropy of an eigenstate satisfying ETH should be proportional to the size of subsystem A, obeying a volume law of scaling.

ETH is not true for all systems, namely, not all systems can thermalize, or act as a reservoir for any of its subsystems. As a result, equilibrium quantum statistical mechanics fails in such kind of systems, namely those that are many-body localized.

1.1.3 Many-Body Localization

MBL is localization with interactions, driven by disorder. As mentioned above, many-body localized system cannot thermalize and fails to satisfy ETH. Due to the breakdown of of equilibrium quantum statistical mechanics in such systems, the MBL transition is invisible to the traditional statistical mechanical ensembles. However, it is an eigenstate phase transition, the dynamics of the system and the properties of the many-body eigenstate change drastically around the transition. The properties that sharply change include: DC conductivity which is finite in ETH phase vanishes in MBL phase; local spectrum changes from continuous to discrete; subsystem entanglement entropy of the eigenstate obeys volume law in the ETH phase but area law in the MBL phase; the spread of entanglement from non-entangled initial condition obeys power law in ETH phase while becomes

logarithmic in MBL phase; dissipation vanishes in MBL phase, and so on. For details of these properties we refer readers to reference[3].

The MBL phase transition occurs when varying disorder strength at a certain value of energy density, or varying energy density when the system is not fully many-body localized (FMBL), where FMBL means that all many-body eigenstates of the system violate ETH thus the mobility edges disappear, this occurs when disorder strength is larger than a critical value. Because we focus on high energy, there is no major difference between Fermion and Boson, but for simplicity and feasibility (of exact diagonalization), we limit the discussion to spin chains, below we list some 1-dimensional quantum spin models which are widely used to study MBL phase transition[1, 16, 17], our work in later chapters is based on the study of the first one.

$$\hat{H} = -\sum_{i=1}^{L-1} J_i \sigma_i^z \sigma_{i+1}^z + J_2 \sum_{i=1}^{L-2} \sigma_i^z \sigma_{i+2}^z + h \sum_{i=1}^{L} \sigma_i^x \tag{1.6a}$$

$$\hat{H} = \sum_{i=1}^{L} \mathbf{S}_i \cdot \mathbf{S}_{i+1} - h_i S_i^z \tag{1.6b}$$

$$\hat{H} = J_1 \sum_{i=1}^{L-1} [(S_i^x S_{i+1}^x + S_i^y S_{i+1}^y) + S_i^z S_{i+1}^z] + \sum_{i}^{L} h_i S_i^z + J_2 \sum_{i=1}^{L-2} (S_i^x S_{i+2}^x + S_i^y S_{i+2}^y) \tag{1.6c}$$

the concepts of l-bits is useful to understand some of the properties which have sharp difference in MBL and ETH phases. When a system of N-local, two-state degrees of freedom $\{\sigma_i\}$ which is referred to as p-bits (p, physics) is in FMBL regime, it's argued that the Hamiltonian can be written in terms of Pauli operators

$\{\tau_i\}$ called l-bit (l, localized) as:

$$\hat{H} = E_0 + \sum_i \tau_i^z + \sum_{ij} J_{ij}\tau_i^z\tau_j^z + \sum_{n=1}^{\infty}\sum_{i,j,\{k\}} K_{i\{k\}j}^{(n)}\tau_i^z\tau_{k_1}^z\cdots\tau_{k_n}^z\tau_j^z \tag{1.7}$$

The typical magnitudes of the interactions J_{ij} and $K_{i\{k\}j}^{(n)}$, and their probabilities of being large fall off exponentially with distance. For weakly interacting systems, the l-bits τ_i should have substantial overlap with the bare p-bits σ_i, thus may be viewed as dressed p-bits, with a dressing that falls off exponentially in real space.

For a generic state, its dynamics can be described as that each l-bit τ_i precesses around its z axis, the precession rate is determined by the interactions between τ_i and all other l-bits other than i, namely, $\{\tau_{j\neq i}\}$. Because $\tau_i^z, \forall i$ are constants of motion, there is no dissipation of the system, but because of the precession, there is dephasing, and entanglement even for two initially non-entangled sites. However, the spread of entanglement is much slower than that in ETH phase. That's because in thermal phase, if there is interaction between A and B, as well as between B and C, then entanglement is produced not only between A and B, B and C, but also between A and C because they are interacted through B, as a result, the spread of entanglement will obey power law. But the same case in FMBL regime will not produce entanglement between A and C, because A interacts with B only through τ_B^z which is conserved over time thus cannot be affected by the interaction between B and C.

The l-bits construction is proved to work with FMBL regime, but whether it's valid for systems with mobility edges separating MBL and ETH eigenstates is still an open question. As a result, the above discussion about the spread of entanglement is not clear for Hamiltonians that have both extended and localized eigenstates.

1.2 Gutzwiller Approximation of the Hubbard Model

1.2.1 Hubbard Model

The Hubbard model, proposed by Hubbard in the early 1960's[18], works as a paradigm of correlated electronic systems. Though looking simple, it strikingly captures the most important aspects of the system. It consists of a hopping term describing the electrons' nearest neighbor hopping and contributing as the kinetic energy in the Hamiltonian, and an interaction term which is the energy cost of two electrons being on the same site, due to the repulsion between them.

$$\hat{H} = -t \sum_{\langle i,j \rangle \sigma} c^{\dagger}_{i\sigma} c_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow} \tag{1.8}$$

where $\langle i, j \rangle$ denotes pairs of nearest neighbor lattice sites. In general further hopping terms or more interactions can be added to give rise to an extended Hubbard model. Two dimensionless parameters characterizing this Hamiltonian are the strength of the Coulomb interaction U/t and the filling fraction n.

1.2.2 Gutzwiller approximation

Though looking simple, the Hubbard model has analytical solutions for the ground state only in one-dimension. To obtain approximate solutions for higher dimension, Gutzwiller developed a variational method called Gutzwiller projection[7–9]. The idea of this approach is to construct a trial wavefunction from a non-interacting one by suppressing the double occupation (see Fig 1.3) which costs energy due to the repulsion interaction. The amount of double occupation weight reduction is determined variationally thus to minimize the ground state energy, so it depends on the value of U/t, the larger the repulsion, the more the double occupation weight is reduced, when $U = 0$, the weight is not reduced at all so it

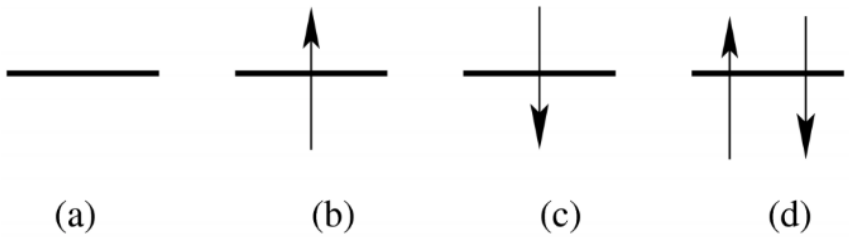

Figure 1.3: Illustration of the local Fock space with zero occupation (empty)(a), single occupation (b, c) and double occupation (d).In a non-interacting wavefunction, the four local states have the same weight as long as the order parameters are fulfilled, whereas the repulsion makes double occupation less favored. Gutzwiller wavefunction reduce the weight of the double occupation state from the non-interacting wavefunction to reflect the role of the interaction.

recovers the non-interacting result.

To formalize, the Gutzwiller wavefunction can be written as:

$$|\Psi\rangle = \hat{P}|\Psi_0\rangle \tag{1.9}$$

where $|\Psi_0\rangle$ is the non-interacting function and $\hat{P}$ is the projection operator which reduces the double occupation on all sites by:

$$\hat{P} = \prod_i \hat{P}_i = \prod_i g_i^{\hat{D}_i} \tag{1.10}$$

$\hat{P}_i$ is the local projection operator acting on each site. $\hat{D}_i = \hat{n}_{i\uparrow}\hat{n}_{i\downarrow}$ is the double occupation projection operator acting on site i, g_i's are variational parameters satisfying $0 \leq g_i \leq 1, \forall i$, the first equality holds when $U \to \infty$ thus the double occupation is completely projected out, while the second one hold when $U = 0$, recovering non-interacting wavefunction.

Under this formula, the expectation value of any operator can be computed

with the projected Gutzwiller wavefunction as:

$$\langle \hat{O} \rangle = \frac{\langle \Psi_0 | \hat{P}\hat{O}\hat{P} | \Psi_0 \rangle}{\langle \Psi_0 | \hat{P}\hat{P} | \Psi_0 \rangle} = \frac{\langle \hat{P}\hat{O}\hat{P} \rangle_0}{\langle \hat{P}^2 \rangle_0} \tag{1.11}$$

where $\langle \cdots \rangle$ denotes the average in non-interacting case. In the same manner, the ground state energy can be approximated as:

$$E(g) = \langle \hat{H} \rangle = \frac{\langle \hat{P}\hat{H}\hat{P} \rangle_0}{\langle \hat{P}^2 \rangle_0} \tag{1.12}$$

the variational parameters g_i's are determined by minimizing the ground state energy $E(g)$, in the homogeneous case, $g_i = g, \forall i$. In the special case of one or infinite dimension, this is analytically solvable[19, 20] and the result agrees well with the exact result.

Before and after the projection, the charge density should remain the same, however, this original form of the projection operator $\hat{P}_i = g_i^{\hat{D}_i}$ does not reflect this constraint. In order to ensure that $n_{i\sigma} = n_{i\sigma}^0$, while the superscript ^{0} indicates the quantity in non-interacting case, two extra fugacities in $\hat{P}_i$ are required. The projection operator is now modified to be:

$$\hat{P}_i = y_{i\uparrow}^{\hat{n}_{i\uparrow}} y_{i\downarrow}^{\hat{n}_{i\downarrow}} g_i^{\hat{D}_i} \tag{1.13}$$

Now Lagrange multipliers can be introduced to ensure that the charge density remains the same after projection: $\lambda_{i\sigma}(n_{i\sigma}^0 - n_{i\sigma})$.

Now the goal is to formalize a renormalized non-interacting Hamiltonian $\hat{H}_{re}$ satisfying $\langle \hat{H}_{re} \rangle_0 = \langle \Psi | \hat{H} | \Psi \rangle$, note that the equality holds only at the level of average value. To facilitate the derivations to achieve this goal, all four local state projection operators are introduced below, namely empty projection $\hat{E}_i$, single occupied σ projection $\hat{Q}_{i\sigma}$ and double occupation projection operator $\hat{D}_i$ which

was already seen above,

$$\hat{D}_i = \hat{n}_{i\uparrow}\hat{n}_{i\downarrow} \tag{1.14a}$$

$$\hat{E}_i = (1-\hat{n}_{i\uparrow})(1-\hat{n}_{i\downarrow}) \tag{1.14b}$$

$$\hat{Q}_{i\uparrow} = \hat{n}_{i\uparrow}(1-\hat{n}_{i\downarrow}) \tag{1.14c}$$

$$\hat{Q}_{i\downarrow} = \hat{n}_{i\downarrow}(1-\hat{n}_{i\uparrow}) \tag{1.14d}$$

The four local states form a complete local Fock space thus

$$\hat{E}_i + \hat{D}_i + \hat{Q}_{i\uparrow} + \hat{Q}_{i\downarrow} = 1 \tag{1.15}$$

The Gutzwiller projection operator can be expanded and if ignoring higher order terms it turns to be:

$$\begin{aligned}\hat{P}_i &= y_{i\uparrow}^{\hat{n}_{i\uparrow}} y_{i\downarrow}^{\hat{n}_{i\downarrow}} (1-(1-g_i)\hat{D}_i) \\ &= y_{i\uparrow}^{\hat{n}_{i\uparrow}} y_{i\downarrow}^{\hat{n}_{i\downarrow}} (\hat{E}_i + \hat{Q}_{i\uparrow} + \hat{Q}_{i\downarrow} + g_i\hat{D}_i) \\ &= \hat{E}_i + y_{i\uparrow}\hat{Q}_{i\uparrow} + y_{i\downarrow}\hat{Q}_{i\downarrow} + g_i y_{i\uparrow} y_{i\downarrow}\hat{D}_i \end{aligned} \tag{1.16}$$

Let $z_i = \langle \hat{P}_i^2 \rangle_0$, after some detailed algebra the hopping term in the Hubbard model can be renormalized as:

$$\begin{aligned}\langle c_{i\sigma}^\dagger c_{j\sigma}\rangle &= \frac{\langle \hat{P}_i c_{i\sigma}^\dagger \hat{P}_i \hat{P}_j c_{j\sigma} \hat{P}_j \rangle_0}{z_i z_j} \\ &= g_{i\sigma}^t g_{j\sigma}^t \langle c_{i\sigma}^\dagger c_{j\sigma}\rangle_0 \end{aligned} \tag{1.17}$$

where we have:

$$g_{i\sigma}^t = \frac{y_{i\sigma}(1-n_{i\bar{\sigma}}^0)}{z_i} + \frac{y_{i\sigma} y_{i\bar{\sigma}}^2 g_i n_{i\bar{\sigma}}^0}{z_i} \tag{1.18}$$

Below we try to express $g_i, y_{i\sigma}, z_i$ in terms of $d_i, e_i, q_{i\sigma}$ and $d_i^0, e_i^0, q_{i\sigma}^0$, using the

relations:

$$d_i = \langle \hat{D}_i \rangle = \frac{g_i^2 y_{i\uparrow}^2 y_{i\downarrow}^2 d_i^0}{z_i} \tag{1.19a}$$

$$e_i = \langle \hat{E}_i \rangle = \frac{e_i^0}{z_i} \tag{1.19b}$$

$$q_{i\sigma} = \langle \hat{Q}_{i\sigma} \rangle = \frac{y_{i\sigma}^2 q_{i\sigma}^0}{z_i} \tag{1.19c}$$

The second equality of each of the above expressions is obtained using 1.14a-1.14d. Then the Gutzwiller coefficient $g_{i\sigma}^t$ in 1.18 can be written as:

$$g_{i\sigma}^t = \sqrt{\frac{e_i q_{i\sigma}}{e_i^0 q_{i\sigma}^0}}(1 - n_{i\bar{\sigma}}^0) + \sqrt{\frac{d_i q_{i\bar{\sigma}}}{d_i^0 q_{i\bar{\sigma}}^0}} n_{i\bar{\sigma}}^0 \tag{1.20}$$

Considering that $q_{i\sigma} + d_i = n_{i\sigma}$, $d_i + e_i + q_{i\uparrow} + q_{i\downarrow} = 1$ and $n_{i\sigma} = n_{i\sigma}^0$, eq. 1.20 can further be written as d_i and $n_{i\sigma}^0$:

$$g_{i\sigma}^t = \sqrt{\frac{(n_{i\sigma}^0 - d_i)(1 + d_i - n_{i\sigma}^0)}{n_{i\sigma}^0(1 - n_{i\sigma}^0)}} + \sqrt{\frac{d_i(n_{i\bar{\sigma}}^0 - d_i)}{n_{i\bar{\sigma}}^0(1 - n_{i\bar{\sigma}}^0)}} \tag{1.21}$$

While when $n_{i\sigma}^0$ is known, for example, in paramagnetic phase, $n_{i\sigma}^0 = n/2$, while in ferromagnetic phase $n_{i\sigma}^0 = n/2 + m$, where n is the electron density and m is the magnetic order, d is the only variational variable to minimize the energy. The renormalized Hamiltonian can be written as:

$$\hat{H} = -t \sum_{\langle i,j \rangle \sigma} g_{i\sigma}^t g_{j\sigma}^t c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i d_i - \mu \sum_{i\sigma} \hat{n}_{i\sigma} \tag{1.22}$$

under constraints $\sum_{i\sigma} \lambda_{i\sigma}(\hat{n}_{i\sigma} - n_{i\sigma}^0)$ which ensure the electron density unchanged before and after the projection.

1.2.3 Mott transition at half filling

Consider the Hubbard model at half filling, naively it may be thought as a metal as the band is half filled, however, as the double occupation costs extra energy, when U is large enough hopping process is not allowed as it creates double occupation inevitably in the half filled system. As a result, there is a metal-insulator transition when varying U, this type of phase transition is called Mott transition, which can be described by Gutzwiller approximation, as shown below. This metal-insulator quantum phase transition is the starting point of Chapter IV where we will generalize the approximation approach to finite temperature.

When there is no magnetism, in the homogeneous half filling case, $n^0_{i\uparrow} = n^0_{i\downarrow} = 1/2$, $d_i = d$, taking them into eq. 1.21, g^t_σ can be calculated as:

$$g^t_\sigma = 4\sqrt{d(1/2 - d)} \tag{1.23}$$

The variational energy per site on a square lattice thus turns to:

$$E(d)/N_s = -8t\langle c^\dagger_{i\sigma} c_{i\sigma}\rangle_0 \cdot 16d(1/2 - d) + U \cdot d \tag{1.24}$$

where N_s is the number of lattice sites. d is determined by minimizing the variational energy $E(d)/N_s$:

$$d = \frac{1}{4} \cdot (1 - \frac{U}{64t \cdot \langle c^\dagger_{i\sigma} c_{i\sigma}\rangle_0}) \tag{1.25}$$

d vanishes at $U_c = 64t \cdot \langle c^\dagger_{i\sigma} c_{i\sigma}\rangle_0 \approx 12.9t$. The value of d and g^t_σ varying with U are shown in Fig. 1.4.

From Fig. 1.4, g^t_σ decreases with the increase of U and vanishes at U_c, meaning that the interaction renormalize the bandwidth down, at the critical point the band is flattened and the effective mass of the electron diverges, which is how

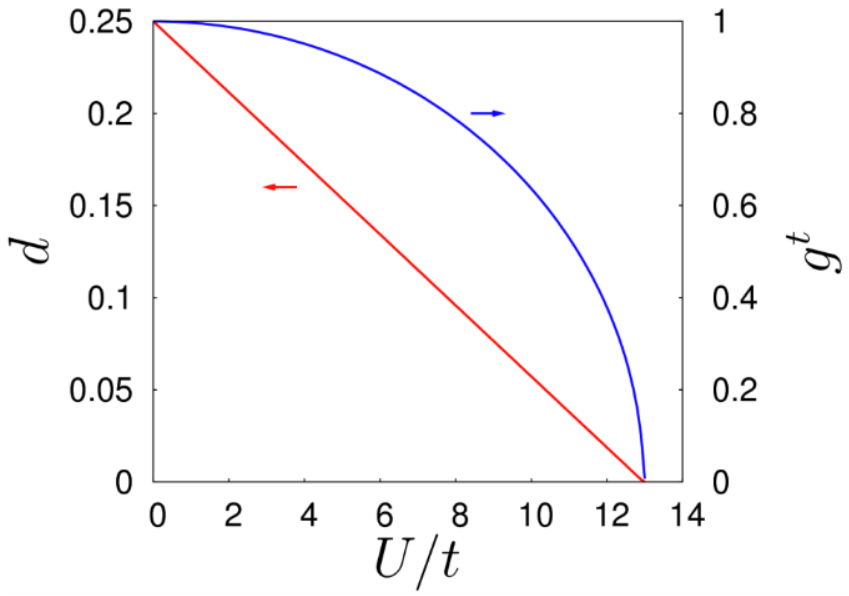

Figure 1.4: Metal-insulator transition at half filling. Double occupancy d and Gutzwiller factor as functions of Hubbard U on a square lattice.

Brinkman and Rice describe this metal-insulator transition.

The Gutzwiller approximation which treats the interaction term as a mean field underestimates the quantum fluctuation. According to the results obtained from more accurate numerical methods such as variational Monte Carlo simulation, such metal-insulator transition will never happen at finite U, indicating that the Gutzwiller approximation overestimate the tendency that the system becomes frozen, however, the idea that the bandwidth is down renormalized by the interaction is valid.

CHAPTER II

Interpretable Machine Learning Study of Many-Body Localization Transition in Disordered Quantum Spin Chain

2.1 Introduction

Many-body localization (MBL) refers to a class of correlated systems that fail to thermalize in the sense that they violate the eigenstate thermalization hypothesis (ETH) [3, 21–23]. As a consequence, certain memories of the local initial conditions can be forever remembered in conserved local observables. They thus have the potential to robustly store quantum information [5]. Compared to the conventional thermal phase, the MBL phase has many novel characteristic properties. The hallmark of the MBL phase is that the eigenstate entanglement entropy follows the area-law instead of the volume-law in the thermal phase [1, 16, 24–31]. The MBL phase has zero DC conductivity [32] and discrete local spectrum [4]. The statistics of the energy level spacing in the MBL phase is described by the Poisson distribution, in contrast to the Wigner-Gaussian distribution typical in the thermal phases [3, 5, 16, 29, 33–35].

The properties of the entanglement entropy and the level spacing have been commonly used to study MBL-thermal phase transition [1, 16, 36–40]. However,

the intrinsic many-body problem makes the study of the critical phenomena very challenging due to the sample size limitations and the nonperturbative nature of strong disorder. Despite the formal analogy to the mobility edge problem in the single particle Anderson localization [41], such basic questions of whether the MBL-ETH transition can be viewed as a localization transition in the many-body Hilbert space remains controversial. It is known that Anderson localization is stable against weak electron-electron interactions, which suggests that the MBL phase would emerge when disorder is strong enough [32]. One of the most profound and powerful physical quantities widely used to identify the Anderson localization transition is the inverse participation ratio (IPR) [42] that measures the (inverse) of the spatial coverage of the single-particle eigenstates. One therefore asks if the MBL arises through the localization of the many-body states in the configurational Hilbert space, and if the scaling behavior of properly generalized IPR can be used to determine the MBL phase transition. Several theoretical studies have shown that the behavior of the IPR (or its inverse) and the entanglement entropy share similarities [43–45] and are directly related in the single particle picture [46], whereas others offer opposite arguments [16, 47]. Recent experimental measurements also explored and demonstrated the connections between Hilbert space localization and energy level statistics[48].

In this work, we apply machine learning to the classification of two different phases, the ETH and the MBL. We will also explore and extract useful information concerning the above questions from a machine learning perspective. Specifically, we build and operate the support vector machine (SVM), designed for the random transverse-field Ising chain. First, we demonstrate that the trained SVM with appropriate kernel choice is able to distinguish the two phases and determine the phase boundary. For our model, we only require training data from two different energy densities to make the trained SVM work for the whole energy

spectrum. This fact ensures that during the training process, the models are built on properties of the MBL phase itself which should not depend on energy. Compared to training and testing at a fixed energy density and repeat the process multiple times in the full energy space to determine the transition line, training only once is much more computation cost-saving, especially considering that it is often expensive to generate class labels. Finally, we try to study and understand how the SVM makes the decision. We find strong evidence that the SVM has the ability to automatically choose a decision function which is very closely related to the many-body IPR defined in the configuration space.

2.2 Model and Method

2.2.1 Transverse-field disordered quantum Ising chain

The quantum transverse-field Ising chain is known to develop the MBL phase when the disorder strength is strong. The Hamiltonian of the system is given by [1]

$$\hat{H} = -\sum_{i=1}^{L-1} J_i \sigma_i^z \sigma_{i+1}^z + J_2 \sum_{i=1}^{L-2} \sigma_i^z \sigma_{i+2}^z + h \sum_{i=1}^{L} \sigma_i^x \tag{2.1}$$

where σ^x and σ^z are Pauli matrices and L is the number of sites in the chain. In Eq. (3.1), the second nearest neighbor coupling J_2 and the transverse eternal field h will be assigned uniform and nonrandom values, whereas the nearest neighbor coupling is site-dependent, $J_i = J + \delta J_i$, where J is a constant and δJ_i is randomly taken from a uniform distribution $[-\delta J, \delta J]$. Thus δJ measures the disorder strength. For a certain disorder realization, the energy E of the many-body eigenstates of H is bounded within a bandwidth $W = E_{max} - E_{min}$. Consider a disordered ensemble of H, the appropriate dimensionless energy is defined by the energy density $\epsilon = 2(E - E_{min})/W$ relative to the total bandwidth, within a small window around ϵ. The density of states of this model at $\delta J = 1.8$ when

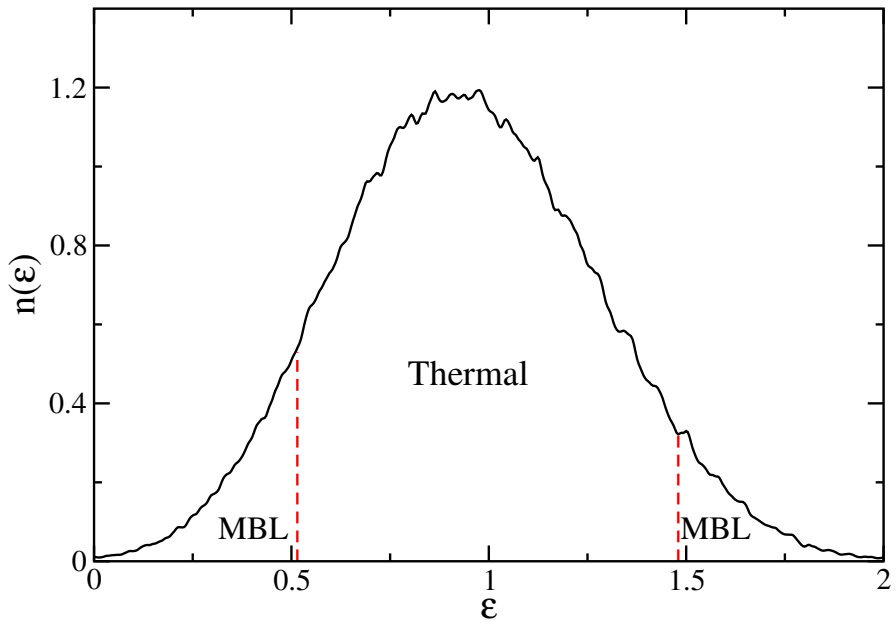

Figure 2.1: Density of state of Hamiltonian in Eq. (3.1) at $\delta J = 1.8$ for a specific disorder configuration. ϵ is the energy density. The mobility edges separating thermal and MBL phases are determined according to supplementary material of [1].

$L = 14$ for a specific disorder configuration is shown in Fig. 2.1. For a given set of J, J_2, and h, the transition between the thermal (ETH) and MBL phases corresponds to a boundary in the phase diagram spanned by δJ and ϵ. Here we set $J_2 = 0.5h = 0.3J$.

We express the many-body quantum states and the Hamiltonian matrix in the spin configuration basis, which is constructed by direct product states of the *local* Hilbert space $\{\sigma_i^z\}$. In addition to being natural, this basis is non-entangled and suitable for introducing the many-body IPR to describe the localization in the spin configuration basis. We work in this basis throughout the rest of the paper.

2.2.2 Data for machine learning

Instead of dividing the system into two subsystems A and B to calculate the reduced density matrix of an eigenstate $\rho_A = \mathrm{Tr}_B|\Psi\rangle\langle\Psi|$ and using the entangle-

ment spectrum as the training data set [49, 50], we directly feed the probability density of the eigenstate $|\Psi\rangle$ computed in the spin basis to the machines as the training data set. The reason for doing so is that, although by preprocessing the training data can reduce the dimension and filter out redundant information, useful information contained in the wavefunction of the entire system can also be lost. Since the entanglement entropy is not the only quantity that can characterize the MBL phase, we thus classify the probability density of the wave function instead of the entanglement spectra. This method not only allows the exploration of other characteristic physical quantities of MBL in the entire system, but also stages a test on the power of machine learning: if only the minimally processed knowledge is provided in the training data, will machine learning be able to find out the relevant physical property to be used for classification by itself?

Our results show that the answer is affirmative. In addition, the algorithm turns out to be remarkably efficient for our model: only input wave functions at two different energy densities are used as the training set and the trained model is able to determine the transition region at all energy densities and the mobility edge for any disorder strength. In other words, by training with wave functions generated at four corner points on the $(\delta J, \epsilon)$-plane, the models are able to produce the complete phase boundary in the 2-parameter phase diagram. It is also remarkable that the SVM is capable of capturing certain generic properties for all energy densities in making the decision, rather than being trapped by energy-specific properties. This part is presented in detail together with the classification results and the decision function detection in Section 2.3.

2.2.3 Support vector machine

There are many machine learning models that are widely used for data classification. Some of them have been applied to study phase transitions in many-body

systems, such as artificial neural networks [49–53], clustering via principal component analysis [54], and kernel method for support vector machine (SVM) [55, 56]. Here we focus on the last one due to its better interpretability.

SVM is one of the most successful model for binary classification, which aims to linearly separate data belonging to two classes $\{+1, -1\}$, making the distance between the separating hyperplane and its nearest data points in both classes as large as possible. In other words, for any hyperplane separating the two classes of data, there exists a region where we can pin the separating hyperplane without changing the accuracy of classification. This region is called the margin and we want to find the hyperplane corresponding to the maximum margin. Fig. 2.2(a) is a schematic plot of how a separation plane separates different phases with largest margin in a two-dimensional feature space.

The hyperplane satisfying this requirement can be described by the linear equation: $\vec{w} \cdot \vec{x} + b = 0$, where $\vec{w}$ is the vector perpendicular to the hyperplane and $\vec{x}$ denotes any point on the hyperplane. Since only the direction of $\vec{w}$ matters, we can rescale the modulus of $\vec{w}$ and make the distance between the separating hyperplane and its closest data points equal to one. Denoting those data points closest to the separating hyperplane as $\vec{x}^{SV}$ (where the superscript SV stands for support vectors), we have, after rescaling, $|\vec{w}\cdot\vec{x}^{SV}+b| = 1$. As a result, the distance from $\vec{x}^{SV}$ to the hyperplane, $|(\vec{x}^{SV} - \vec{x}) \cdot \frac{\vec{w}}{|\vec{w}|}| = \frac{1}{|\vec{w}|}$ is what we want to maximize. Equivalently, we can minimize $\frac{1}{2}\vec{w} \cdot \vec{w}$ subject to the condition $y_n(\vec{w} \cdot \vec{x}_n + b) \geq 1$, where $\vec{x}_n$ is any of the training data samples in the two classes $y_n = \pm 1$, because the distance from any of them to the separating hyperplane is at least 1.

Next, consider the case where the data points are not completely linearly separable, i.e. a few of them would fall into the margin of the linear-separating hyperplane. As a result, the above constraint can be adjusted according to $y_n(\vec{w}^T \cdot \vec{x}_n + b) \geq 1 - \xi_n$, where $\xi_n \geq 0$ for all data points and the total vio-

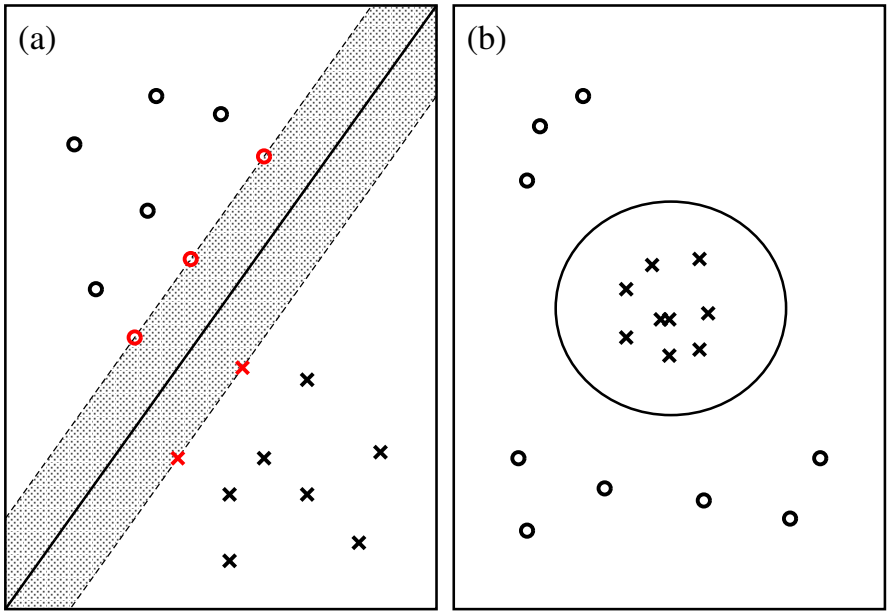

Figure 2.2: (a) A separating plane (solid line) separates two different phases (labeled as circles and crosses respectively) with the largest margin (shaded area) in the 2-dimensional feature space. The red circles and crosses mark the support vectors that are closest to the separating plane. (b) The large circle in the original 2-dimensional feature space is a separating hyperplane in higher dimensional space after the transformation. Such a transformation makes data points that are not linearly separable in its original space linearly separable in the transformed higher dimensional space.

lation is the sum of all ξ_n. Using the Kuhn-Tucher theorem, the minimization of $\frac{1}{2}\vec{w}\cdot\vec{w}$ under the constraints can be achieved by minimizing the following effective Lagrangian,

$$\begin{aligned}\mathcal{L}(\vec{w},b,\vec{\xi},\vec{\alpha},\vec{\beta}) = \frac{1}{2}\vec{w}^T\cdot\vec{w} + C\sum_{n=1}^{N}\xi_n \\ -\sum_{n=1}^{N}\alpha_n[y_n(\vec{w}^T\cdot\vec{x}_n + b) \\ -(1-\xi_n)] - \sum_{n=1}^{N}\beta_n\xi_n \end{aligned} \tag{2.2}$$

where N is the total number of training samples and $\alpha_n, \beta_n \geq 0$ are the Lagrangian multipliers enforcing the constraints. The second term on the r.h.s of Eq. (2.2) is the regularization term that specifies the price that violations of the margin have to pay. Increasing C means less tolerance for violating the margin, thus yields more complex models, whereas decreasing C makes the price of violation smaller, thus avoids overly fitting the noise. The "hyperparameter" C should be determined by grid search in a manually specified subset of values. We take the value of C that leads to the best validation result. The validation data samples are generated at values of δJ in the same range as the testing data set, but for different disorder realizations.

Minimizing $\mathcal{L}$ with respect to $\vec{w}$, b, and ξ_n first leads to,

$$\vec{\nabla}_{\vec{w}}\mathcal{L} = \vec{w} - \sum_{n=1}^{N}\alpha_n y_n \vec{x}_n = 0 \tag{2.3}$$

$$\frac{\partial\mathcal{L}}{\partial b} = -\sum_{n=1}^{N}\alpha_n y_n = 0 \tag{2.4}$$

$$\nabla_{\xi_n}\mathcal{L} = C - \alpha_n - \beta_n = 0 \tag{2.5}$$

Plugging Eqs (2.3-2.5) into Eq. (2.2), we can get rid of the variables $\vec{w}$, $\vec{\xi}$, and b,

and obtain $-\mathcal{L}(\vec{\alpha})$ which is to be minimized with respect to $\vec{\alpha}$:

$$\frac{1}{2}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix}^T\begin{bmatrix}y_1y_1\mathcal{K}_{11} & y_1y_2\mathcal{K}_{12} & \dots & y_1y_N\mathcal{K}_{1N}\\ y_2y_1\mathcal{K}_{21} & y_2y_2\mathcal{K}_{22} & \dots & y_2y_N\mathcal{K}_{2N}\\ \vdots & \vdots & \ddots & \vdots\\ y_Ny_1\mathcal{K}_{N1} & y_Ny_2\mathcal{K}_{N2} & \dots & y_Ny_N\mathcal{K}_{NN}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix} - \vec{1}\cdot\vec{\alpha} = -\mathcal{L}(\vec{\alpha}) \tag{2.6}$$

under the constraints that $\sum_{n=1}^{N}\alpha_n y_n = 0$ and $0 \le \alpha_n \le C, \forall n$, where the $\mathcal{K}_{ij} = \vec{x}_i \cdot \vec{x}_j$ are called the kernel. Note that only a few (out of N) of the α_n are nonzero, otherwise there is a high risk for over-fitting. Those nonzero α_n correspond to the data points that are closest to the separating hyperplane. They are the so-called called support vectors because they are what determine the separating hyperplane in the end. After obtaining the α_n, $\vec{w}$ can be obtained from Eq. (2.3) by

$$\vec{w} = \sum_{k=1}^{N_{SV}} \alpha_k^{SV} y_k^{SV} \vec{x}_k^{SV} \tag{2.7}$$

where $\vec{x}_k^{SV}$ is one of the N_{SV} number of the support vectors.

In the above linear algorithm, the kernel $\mathcal{K}_{ij}$ is simply the inner product of two data points $\vec{x}_i$ and $\vec{x}_j$. However, in most of the realistic cases, the data sets are not linearly separable and we have to transform a data point from a vector $\vec{x}$ in its original space X to a vector $\vec{z}$ in a higher dimensional space Z. Fig.2.2(b) illustrates a simple example of such kind of transformation. If the original X space is 2-dimensional and represented by (x_1, x_2), the simplest transformation to the higher dimensional space $X \to Z$ corresponds to be $(x_1, x_2) \to (x_1^2, \sqrt{2}x_1x_2, x_2^2)$. Consequently, the kernel in Z space is $\mathcal{K}_{ij} = \vec{z}_i \cdot \vec{z}_j = (\vec{x}_i \cdot \vec{x}_j)^2$. In the actual calculations, we only need to know the values of the kernel in order to minimize Eq. (2.6) to obtain α_n and thus the decision function. In fact, a set of input data

can be raised to any order by choosing the general form of the polynomial kernel $\mathcal{K}_{ij} = \mathcal{K}(\vec{x}_i, \vec{x}_j) = (c_0 + \gamma \vec{x}_i \cdot \vec{x}_j)^d$, or even transformed to infinitely dimensions of space by choosing a radial basis function (RBF) kernel $\mathcal{K}_{ij} = exp(-\gamma |\vec{x}_i - \vec{x}_j|^2)$. The resulting decision function is determined by the value of the kernels according to:

$$f(\vec{x}) = \text{sign}\left(\sum_{k=1}^{N_{SV}} \alpha_k^{SV} y_k^{SV} \mathcal{K}(\vec{x}_k^{SV}, \vec{x}) + b\right) \tag{2.8}$$

where $\vec{x}_k^{SV}$'s are support vectors.

2.3 Phase Classification and Decision Function

2.3.1 Classification result and phase diagram

In our case, both the training and testing data sets are composed of probability density of the eigenstate wavefunctions of the Hamiltonian in Eq. (3.1) obtained by exact diagonalization, labeled as MBL $(+1)$ or ETH(-1). We choose $\delta J = 0.15 \pm 0.05$ and energy densities $\epsilon = 59/60$ and $\epsilon = 19/60$ which are deep in ETH phase and $\delta J = 9.0 \pm 1.0$ at the same energy densities which are deep in MBL phase to generate 18000 wavefunctions, 4500 for each set of $(\delta J, \epsilon)$, and use their probability densities as the training set. We will demonstrate that by training the machine learning models at two different energy densities, the precise values of which are not important, we can obtain a model that works for determining the phase diagram in the whole energy spectrum. More detailed discussion and the possible implications of this remarkable finding will be given at the end of this subsection.

We first train the SVM with different kernels, including the linear kernel, the polynomial kernel with $d = 2, 3, 4, 5, 6$, and the RBF kernels. Since we only wish to keep the homogeneous terms, we choose $c_0 = 0$ in $\mathcal{K}_{ij} = (c_0 + \gamma \vec{x}_i \cdot \vec{x}_j)^d$ for the polynomial kernels. By grid-search we find that in this case the models are

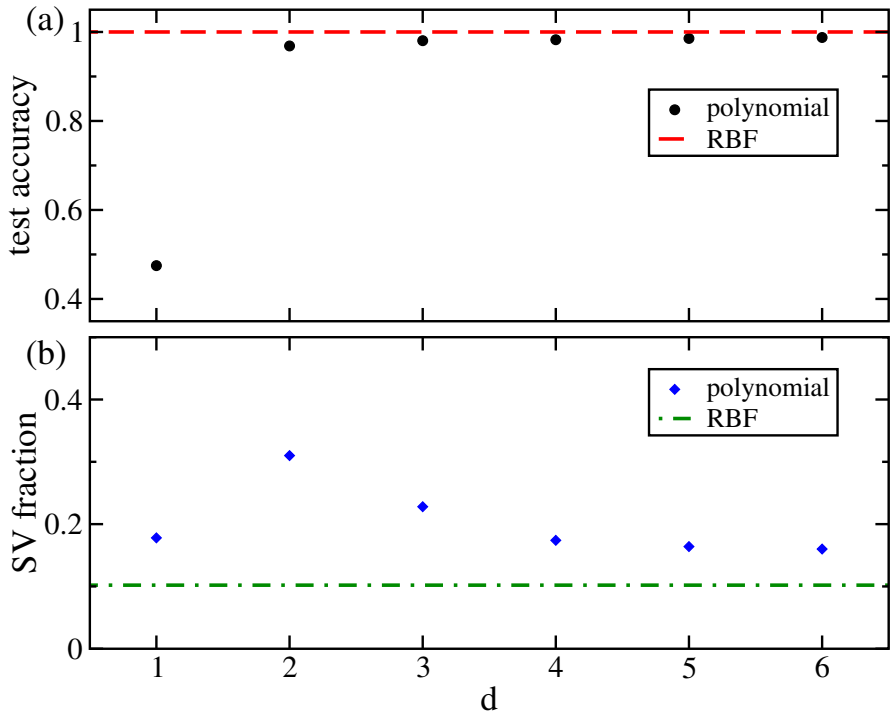

Figure 2.3: (a) The test accuracy as a function of the order (d) of the polynomial kernel. The black dots denote the test accuracy. It increases from 47.5% for the linear kernel at $d = 1$ and approaches 100% corresponding to that of the RBF kernel (the red dashed line). (b) The fraction of support vectors among all training data versus kernel order d shown in the blue squares. The green dash-dot line corresponds to the fraction of SV in the RBF kernel.

not very sensitive to the regularization. Specifically, when C in Eq. (2) is swept through $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$, we find that the test accuracy in the validation set always stays above 96% for all polynomial kernels when $C \in [10^{-2}, 10^2]$, which is unaffected by the order d (excluding the special case $d = 1$, i.e. the linear kernel). Therefore, we choose $C = 1.0$ for our models. For models with polynomial kernels, there exits a threshold of γ in the kernel expressions $\mathcal{K}_{ij} = (c_0 + \gamma \vec{x}_i \cdot \vec{x}_j)^d$ and $\mathcal{K}_{ij} = exp(-\gamma |\vec{x}_i - \vec{x}_j|^2)$, above which the validation accuracy reaches its maximum. We choose $\gamma = 400$, which is large enough to give the optimum validation result for the polynomial models. While for the RBF kernel, we choose $\gamma = 1/2^{L+6}$, which is also determined by validation.

Next, we make a model selection of the kernels to adopt based on their performances on the testing set, then use the selected kernel to proceed with the phase classification. The testing set consists of probability density of wavefunctions generated at $\delta J \in [0.05, 0.45]$ labeled as ETH and $\delta J \in [9.0, 12.0]$ labeled as MBL at $\epsilon = 59/60, 43/60, 31/60, 19/60$. The result for the model selection with $L = 12$ is shown in Fig. 2.3(a). We find the test accuracy in the test set is below 50% for the linear kernel, implying that the linear SVM is unable to distinguish between the ETH and MBL phases. The polynomial SVMs, on the otherhand, all have test accuracy above 96%, meaning that the polynomial SVMs are all qualified phase classifiers. The test accuracy increases with increasing d until reaching about 100% for the RBF kernel.

In Fig. 2.3(b) we show the fraction of support vectors (SV), namely, the number of nonzero α_n among all training data for $L = 12$. The fraction of SV is always smaller than 1/3. Because the number of SV is directly related to the effective degrees of freedom of the model, this indicates that we are not at the risk of over-fitting. In addition, the fraction of SV decreases with increasing d when $d \geq 2$, until it reaches 10.2% for the RBF kernel. Considering that SV are the

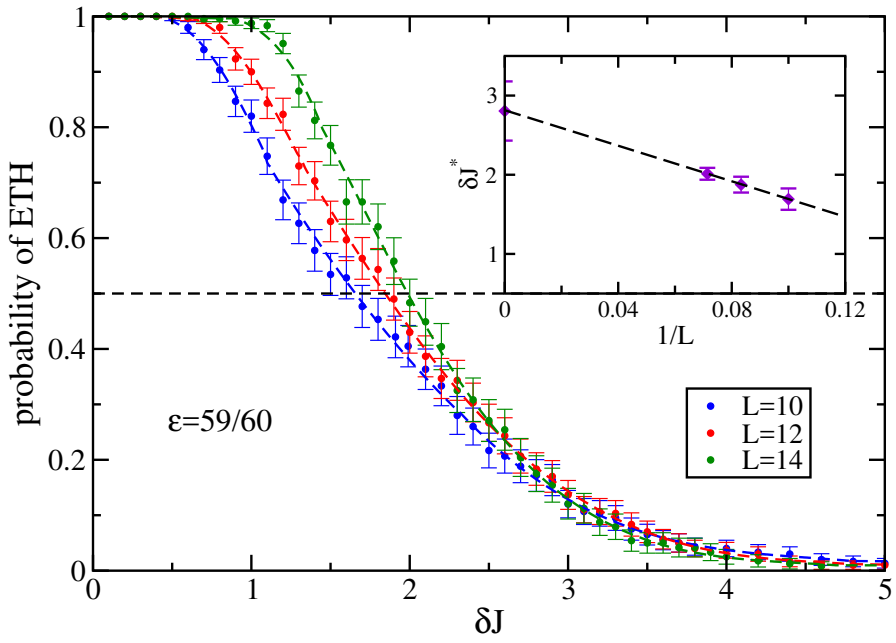

Figure 2.4: The probability that eigen-wavefunction corresponding to energy density $\epsilon = 59/60$ generated at a given δJ is ETH phase for $\delta J \in [0, 5]$. The probability is estimated using the fraction of ETH phase in an ensemble of 300 disorder realizations at energy density $\epsilon = 59/60$ for $L = 10$ (blue dots), $L = 12$ (red dots) and $L = 14$ (red dots) predicted by SVM with RBF kernel. For each size, we take the δJ corresponding to 50% probability of being ETH to be the phase boundary and denote it by δJ^*. The inset shows the finite-size extrapolation of δJ^*. The intercept is interpreted as the phase boundary δJ_c in the thermodynamic limit.

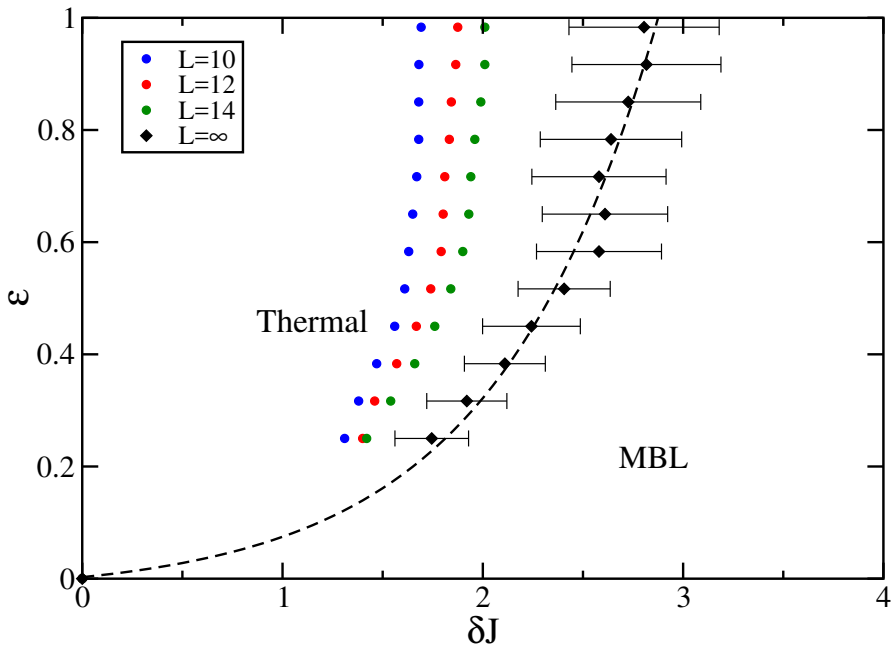

Figure 2.5: Phase diagram of the disordered quantum Ising chain defined in Eq. (3.1) obtained by SVM with RBF kernel. The training was performed for $0.1 \leq \delta J \leq 0.2$ at two energy densities $\epsilon = 59/60$ and $19/60$ labeled as ETH and for $8.0 \leq \delta J \leq 10.0$ at the same two energy densities labeled as MBL. The black diamonds are the critical disorder strengths δJ_c extracted from the large L extrapolations of the finite size transition points (blue, red and green dots) at the different ϵ. The black dashed line is an exponential fit to the phase boundary.

data points most difficult to classify, this result again implies that the SVM with the RBF kernel may be the best choice of model for this study. For $L = 10$ and $L = 14$, the test accuracy versus the order of the polynomial kernel has the same trend as that in $L = 12$ case. Thus, we choose the RBF kernel that gives the best test accuracy (99.81% for $L = 10$ and $\sim 100\%$ for $L = 14$) to search for the phase boundary.

Finally, we use the trained SVM to determine the transition point at different energy densities. For better comparison with previous result [1], we choose $\epsilon = (11+4i)/60, i = 1, 2, \cdots, 12$. For each of the ϵ, we study a series of δJ in the range $[0, 5]$, and for each δJ we consider an ensemble of probability density of eigenstate wavefunctions generated with different disorder realizations/configurations. We

input all eigenstates in an ensemble and compute the fraction of the ETH outputs. When the ensemble is large enough, this fraction corresponds to the probability that a wavefunction generated at the given δJ is in ETH phase. The standard deviation of the probability is calculated according to the central limit theorem. The probabilities are plotted with error bar in Fig. 4 as a function of δJ for different system sizes at a fixed energy density $\epsilon = 59/60$. The probability of being in ETH phase behaves like a soft step function. When δJ is small, namely deep in ETH phase, it approaches 1 because the actual phase should be ETH, whereas for δJ large, i.e. deep in the MBL phase, it approaches 0. In the transition region between the two limiting phases, the probability of being ETH decreases from 1 to 0. We choose the δJ corresponding to ETH probability = 0.5 as the transition point δJ^* (Fig. 2.4) for a given system size L, because it's the disorder strength at which the wavefunctions have half probability to be in ETH phase and half to be in MBL phase thus quantities (like entanglement entropy) that behave differently in these two phases will have the largest standard deviation[1]. As shown in Fig 2.4, with increasing system size, the soft step function becomes steeper, implying that it behaves like a step function in thermal dynamic limit. We regard any disorder strength at which the probability of ETH reaches 0.5 within error as being in transition region, thus to determine the error of δJ^*. As can be seen from Fig. 4, δJ^* exhibits significant size dependence for $L = 14, 12, 10$. In the inset of Fig. 2.4, δJ^* is plotted against $1/L$ and a finite extrapolation within the error bars to the large L limit produces an asymptotic estimate of the δJ_c separating the ETH and MBL phases at this energy density. Repeating this procedure, we computed δJ_c at different energy densities $\epsilon = (11 + 4i)/60, i = 1, 2, \cdots, 12$ shown in the phase diagram Fig. 2.5. The phase boundary separating the ETH and the MBL phases is obtained by an exponential fit to the data, which qualitatively agrees with the result obtained from scaling the variance of the entanglement entropy [1].

It is important to note that the phase diagram cannot be obtained had the data at only one energy density been used as the training set. Indeed, we started off training the model at a single energy density ($\epsilon = 59/60$ or $19/60$) and tested the ability of the model to determine the phase boundaries at different energy densities. Surprisingly, the obtained results were quite poor. The testing accuracy in the best case was below 95%. The resulting transition boundary does not vary much with energy and deviates significantly from the one obtained by scaling the variance of the entanglement entropy [1]. This finding is unexpected and remarkable, since it suggests that the information learned by the SVM is controlled by both the energy density and the disorder strength. In order to correctly determine the phase boundary in the two-parameter phase space, the SVM needs to learn to decipher that the information encoded in the wavefunctions come from a two-parameter support in order to avoid being misled by those at different energies. There are at least two possible origins for this novel behavior: (1) this is due to the specifics of the SVM learning algorithm. However, it is worth noting that we find the same property using the neural networks model, which is discussed in detail in the appendix, suggesting that this finding is not specific to a particular machine learning model. It could still arise from the fact that the input to the models, both the training and the processed information, is the probability density of the many-body wavefunctions. (2) An alternative and physically more interesting possibility is that the thermal to MBL transition driven by disorder δJ and the energy density ϵ (mobility-edge like) have different critical properties, such that the training along one direction of the phase diagram (at fixed energy density) doesn't enable the model to learn the transition along the other (at fixed disorder). This is reminiscent of the situation where there are two relevant scaling directions at a critical point. Clearly, more works in the future are needed to fully understand this remarkable property.

2.3.2 Decision function in SVM

As can be seen in Fig. 2.3, the linear SVM completely fails to distinguish between the two phases, resulting in 47.5% test accuracy, in contrast to the worst case of 96.8% for polynomial kernels. We next study the details in the $L = 12$ case in order to corroborate our conclusion that the SVM cannot separate the input data labeled by the two different phases in their original space, and that the phase classification requires the transformation of the inputs to higher dimensional spaces. Fig. 2.6 shows that when using linear kernel the test accuracy is around or below 50% in different trials, even with increasing number of training samples. The origin of this can be traced back to the fact that the probability amplitudes of the wavefunctions are normalized so that the sum of elements in an $\vec{x}$, whether they are from the ETH or the MBL regions, is unity. Thus, one can imagine a $2^L - 1$ dimensional hyperplane in the feature space where all data samples are distributed because of the constraint. The data points corresponding to MBL phase are more likely to be near the edges of that hyperplane, while the ETH data are more likely to be in the center. It is thus impossible to find a hyperplane of the same dimension to separate them. So we have to turn to at least a quadratic kernel. As shown in Fig. 2.6, using a quadratic kernel dramatically increases the test accuracy to at least 91.7% with 10000 training samples, which can be systematically improved further by enlarging the training set. This is what we expect since more training data will reduce model variance, thus improving the test performance.

The unique advantage of the SVM is that one can uncover the exact form of the decision function, although it can be very cumbersome in higher order polynomial kernels and infeasible in the RBF kernel. In the following, we shall limit ourselves to case of the SVM with the quadratic kernel, where the decision function can be written as:

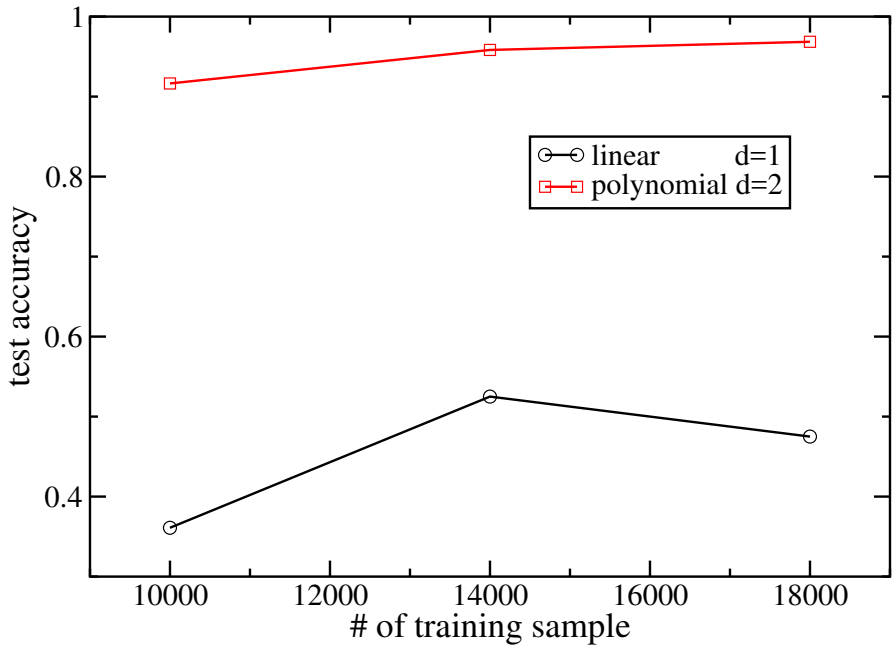

Figure 2.6: The test accuracy obtained from 4500 testing samples of $L = 12$ for the SVM machines using the linear kernel (black line) with $\mathcal{K}(\vec{x_i}, \vec{x_j}) = \vec{x_i} \cdot \vec{x_j}$ and the quadratic polynomial kernel (red line) with $\mathcal{K}(\vec{x_i}, \vec{x_j}) = (\vec{x_i} \cdot \vec{x_j})^2$. The number of training samples used is indicated by the horizontal axis.

$$f(\vec{z}) = \text{sign}\left(\vec{w'}^T \cdot \vec{z} + b\right) = \text{sign}\left(\sum_{i\leq j} w'_{ij} z_{ij} + b\right) \tag{2.9}$$

where $i, j = 1, 2, \cdots, \dim(\mathcal{H})$, and w'_{ij} is each element of $\vec{w'}$ that is coupled to the transformed inputs $\vec{z}$ in the quadratic Z space, according to Eq. (2.7) it can be calculated as:

$$w'_{ij} = \sum_{k=1}^{N_{SV}} \alpha_k^{SV} y_k^{SV} z_{kij}^{SV} = \sum_{k=1}^{N_{SV}} \alpha_k^{SV} y_k^{SV} u_{ij} x_{ki}^{SV} x_{kj}^{SV} \tag{2.10}$$

given the exact form of the transformation from the original X space to the quadratic Z space: $z_{ij} = u_{ij} x_i x_j$ where $u_{ij} = 1$ if $i = j$ and $\sqrt{2}$ if $i < j$. In the same manner, the decision function in Eq. (2.9) can be written in terms of the original basis as:

$$f(\vec{z}) = \text{sign}\left(\sum_{i\leq j} w_{ij} x_i x_j + b\right) \tag{2.11}$$

where $w_{ij} = u_{ij} w'_{ij}$.

In Fig. 2.7, we plot the distributions of the off-diagonal and the diagonal values of $w_{ij}(i < j)$ and w_{ii} for $L = 12$ where $i, j = 1, \cdots, 2^{12}$. Clearly, the distributions of w_{ii} and $w_{ij}(i < j)$ are drastically different. We find that w_{ii} coupling to x_i^2 are positive for all i, with an average of 22.15, which dominates in the decision function over the contributions from $w_{ij}(i < j)$, which can be either positive or negative but are clustered around much smaller magnitudes with an average of $-1.8 * 10^{-3}$. As a result, only the diagonal terms of the kind $x_i^2 = |\langle\{\sigma_i^z\}|\Psi_n\rangle|^4$ contribute essentially to determining the phase region, whereas the cross term of the form $x_i x_j = |\langle\{\sigma_i^z\}|\Psi_n\rangle|^2 \times |\langle\{\sigma_j^z\}|\Psi_n\rangle|^2$ $(i < j)$ do not affect the decision qualitatively. This immediately reminds one of the inverse participation ratio (IPR) that plays a crucial role in the study of the single-particle Anderson localization in disordered media. The generalized definition of the IPR in Fock space of a many-body system

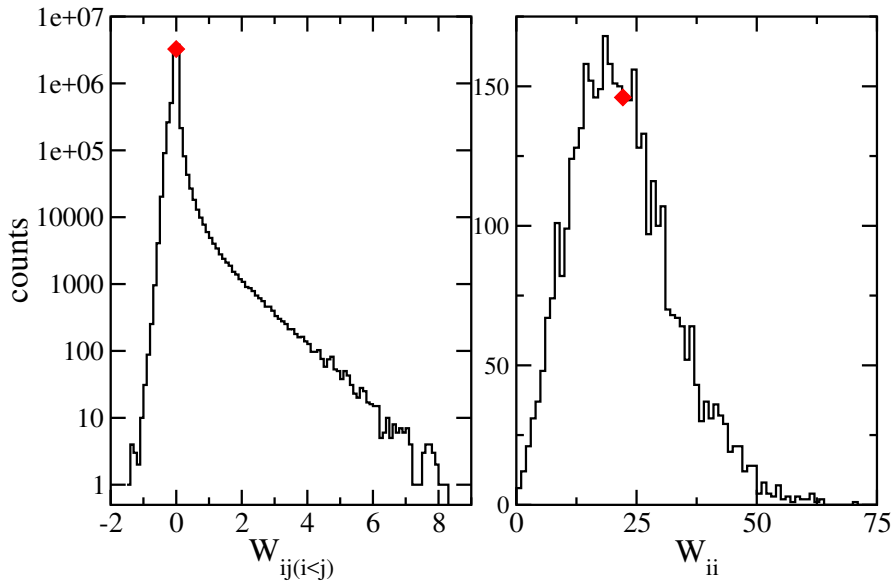

Figure 2.7: Distributions of W_{ij} (left) and W_{ii} (right) for $L = 12$. When $i \neq j$, W_{ij} can be positive or negative, but cluster very close to zero with 96.7% of them distributed in the range $[-0.2, 0.2]$ for an average of $-1.8 * 10^{-3}$ as denoted by the red diamond shown in the left panel. In contrast, the diagonal W_{ii} are much larger. 88.6% of all W_{ii} are larger than 10 with an average of 22.25 as denoted by the red diamond in the right panel.

is:

$$I_q(E_n) = \sum_i |\langle \{\sigma_i^z\} | \Psi_n \rangle|^{2q} \tag{2.12}$$

with $q = 2$. It can also be seen from Fig. 2.7 that most of the w_{ii} are of the same order, indicating that $\langle \{\sigma_i^z\} | \Psi_n \rangle|^4$ for each i contributes almost equally, thus further corroborating that it's a quantity similar to the IPR that acts as the threshold in the decision function of the SVM with the quadratic kernel.

The above analysis and discussion suggest that the decision function of the quadratic SVM is closely related to the many-body IPR $I_{q=2}$. One may wonder if the *total* off-diagonal contribution which after averaging over i is -3.91, is still negligible compared to the diagonal contribution x_i^2 with an average over i being 22.15. A related question is whether the SVM with higher order polynomial kernels also uses decision functions related to the higher order I_q, i.e. if terms

like $|\langle\{\sigma_i^z\}|\Psi_n\rangle|^{2q}$ still dominate in the classification for higher q. Indeed, Fig. 2.3 showed that higher order polynomial kernels lead to better test performance and the test accuracy reaches its maximum for the RBF kernel. It will be instructive to find out the reason for this increase. Is it because the cross terms $x_i x_j$, $i < j$ become more important or more irrelevant, or is it simply because higher order terms are sharper classifiers?

Unfortunately for higher order polynomial kernels, the decision function has poor visualization and becomes even inaccessible in the RBF kernel. So instead of studying the decision functions directly, we preprocess the training data by manually raising each element in the input vector to higher order, removing the cross terms by keeping only terms like $x_i^q = |\langle\{\sigma_i^z\}|\Psi_n\rangle|^{2q}$. Then, we train the linear SVM on the preprocessed data. The test accuracy in the testing set obtained is 99.90% for $q = 2$, 99.75% for $q = 3$ and 99.69% for $q = 4$, suggesting that to correctly distinguish between the MBL and ETH phases, the information from the cross terms are unimportant. Because the test accuracy doesn't change much when varying q in the inputs $|\langle\{\sigma_i^z\}|\Psi_n\rangle|^{2q}$, the IPR of any order equal to or larger than 2 can characterize the phase transition. This result also provides a possible explanation for the increase of test accuracy in the higher order polynomial kernels. The contribution from the cross terms to the decision function may be further suppressed in the higher order polynomial and RBF kernels, which causes the test accuracy to approach that obtained without the cross terms.

To gain further insights, we also applied the three linear SVMs trained on the preprocessed data with $q = 2, 3, 4$ to classifying the data in transition region. The results are shown in Fig. 2.8 at energy density $\epsilon = 59/60$ and $L = 12$. The decision boundary obtained in each case corresponds to $\delta J^* = 1.85 \pm 0.62$, 1.89 ± 0.65 and 1.95 ± 0.70 respectively (shown in colored lines), which agrees well with the result $\delta J^* = 1.88 \pm 0.47$ for the RBF kernel on original data set (shown in black dashed

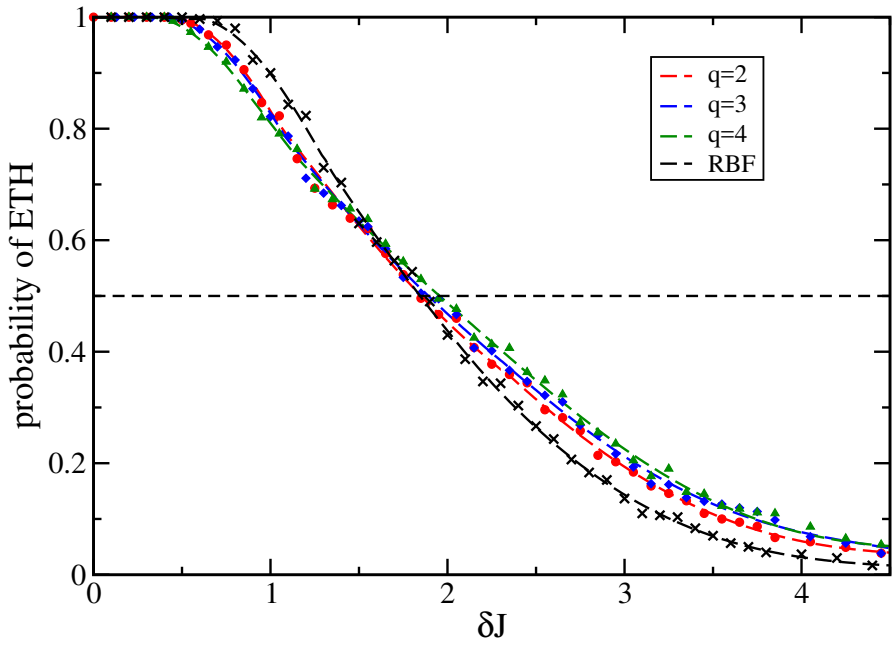

Figure 2.8: Fraction of data points classified as in the ETH phase in an ensemble versus the disorder strength at $\epsilon = 59/60$ and $L = 12$. The colored symbols and dashed lines denote results obtained by the linear SVM trained on the probability density to the q-th power, namely $(x_1^q, x_2^q, \cdots, x_{2^L}^q)$, where $x_i = |\langle \sigma_i^z | \Psi_n \rangle|^2$. Black dashed line is obtained using the RBF kernel trained on the original data set.

line). This further supports our conjecture that when the SVMs with polynomial and RBF kernels search for the decision function, they learn to ignore to a large extent the unnecessary cross terms. As before, the decision function of the linear SVM trained on the preprocessed data has contributions from evenly distributed components, $|\langle\{\sigma_i^z\}|\Psi_n\rangle|^{2q}$ with $q = 2$, of the same order of magnitude. This is consistent with the decision functions being closely related to the IPR in the spin configuration space.

2.3.3 Inverse participation ratio and MBL

The concept of MBL originates from the inability of many-body eigenstates to thermalize in strongly disordered systems. As such, the entanglement entropy S_E between the subsystems has been the common tool used to separate the ETH phase for weak disorder where S_E obeys the volume-law from a MBL phase at for strong disorder where S_E obeys the area-law and the eigenstates fail to thermalize. There remains under investigation, however, an outstanding issue with important physical implications, i.e. if and how MBL is related to the localization of the eigenstates in the many-body Hilbert-space of the entire system under strong disorder and correlation [43–45].

Our interpretable machine learning results described above have shown that, at least for the disordered quantum spin chain studied, the decision function used by the SVM is related to the generalized many-body IPR in Hilbert space. It is known that relating MBL to the localization in Hilbert space requires a choice of basis and is basis dependent. Because we choose the spin configurations as the basis of the Hilbert space, our SVM approach and its consequent interpretability in terms of IPR is also specific to this basis. Furthermore, the SVMs can produce the boundary between the ETH and MBL phases, which is in good agreement with the one obtained by scaling the variance of the entanglement entropy[1], sug-

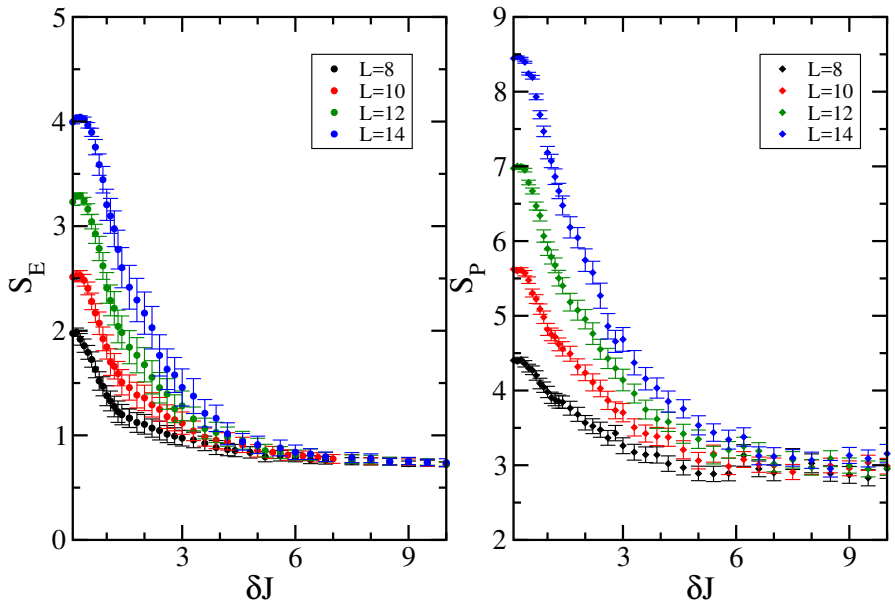

Figure 2.9: The ensemble averaged half-chain entanglement entropy S_E (left panel) and the participation entropy S_P (right panel) plotted versus disorder strength δJ for different length L of the quantum Ising chain at energy density $\epsilon = 59/60$.

gesting that the IPR may have the ability to identify the MBL phase transition as a localization phenomenon in the many-body Hilbert space. In single particle picture, the entanglement entropy defined using the site occupation number basis is deterministically related to the IPR and its multifractal spectrum at the Anderson localization transition point [46]. Unfortunately, it has not been possible to establish the connection between these two quantities for the many-body eigenstates in disordered interacting systems. Motivated by our machine learning results, in the following, we explore the similarities in the behavior of these two quantities in the disordered quantum spin chain.

The entanglement entropy between two partitions separated at the midpoint of the chain is given by, $\mathrm{S_E} = -\mathrm{Tr_L æ_L ln æ_L}$, where $æ_\mathrm{L}$ is the reduced density matrix $\rho_\mathrm{L} = \mathrm{Tr_R}|\Psi\rangle\langle\Psi|$ and L and R denote the left or right half of the chain. In the ETH phase, S_E is an extensive quantity with values proportional to the volume of the subsystem (length $L/2$ of the left half of the chain here) because degrees of

freedom in the subsystems are highly entangled. In the MBL phase, however, the entanglement is limited to the boundary between the subsystems such that S_E is proportional to the boundary area. In 1D systems, it is bounded by a constant.

In order to facility a direct comparison to the entanglement entropy S_E, we convert the IPR into the participation entropy defined by $S_P = -\ln(\sum_i |\langle\{\sigma_i^z\}|\Psi\rangle|^4)$ over the entire system of length L. S_P is commonly used to study the single-particle Anderson localization [57**?**]. In the single-particle case, when the system is in the delocalized phase, S_P is proportional to the logarithm of the size of configuration space and hence the number of lattice sites in the single-particle picture. In the localized phase, on the other hand, S_P is bounded by a constant. At the mobility edge, i.e. the critical point of the metal-insulator transition, S_P exhibits multifractal behavior. For our interacting Ising chain, the size of the configuration space equals 2^L. It is thus natural to expect [16, 58] S_P to be proportional to the length of the chain L up to certain sub-leading terms in the ETH phase, resembling the volume law behavior of the entanglement entropy S_E in the ETH phase. In the MBL phase, it remains to be explored whether S_P is bounded by a constant, namely, whether there exists a genuine localization in the many-body Hilbert space. We calculate both the entanglement entropy S_E and the participation entropy S_P by exact diagonalization at energy density $\epsilon = 59/60$, averaging over ensembles at varying disorder strength δJ. Fig. 2.9 displays the ensemble averaged S_E (left) and S_P (right) as a function of δJ for different length of the chain at $L = 8, \ldots, 14$. There are indeed remarkable similarities in their behaviors. At small δJ, both S_E and S_P exhibit clear linear size (L) dependence characteristic of the volume-law in the ETH phase. As the δJ increases, both S_E and S_P decrease, as does their dependence on the system sizes. In the regime of strong disorder with δJ, the entanglement entropy S_E shows essentially no size dependence, characteristic of a MBL phase with the area-law in 1D. The participation entropy S_P

also displays a much reduced size-dependence, which disappears for the largest sample sizes $L = 10$ and 12 at large disorder δJ. While a definitive conclusions would require numerical studies of even larger system sizes which are beyond our current size limit, these results together with those from the interpretable machine learning studies bring sufficient new insights and raise the possibility of studying theoretically as well as experimentally[48] other physical quantities more directly connected to the localization of the many-body eigenstates in the Hilbert space.

2.4 Summary and Conclusions

We presented in this paper an interpretable machine learning classification of the thermal and MBL phases in a disordered quantum Ising spin chain. Specifically, the SVMs were built with different types of kernels of the probability density of the exact eigenstate wavefunctions. We find that training the machines with data at a minimal of two different energy densities and two disorder strengths corresponding to the limiting cases deep in the thermal and MBL phases, the SVMs are able to classify the phases in the *entire* transition regime and determine the boundary separating the two phases at *all* energy densities. The phase boundary determined by machine learning is in good qualitative agreement with that obtained by scaling the variance of the entanglement entropy[1]. These results show that the decision function of the SVM is a general two-parameter quantity, i.e. the energy density and disorder strength, capable of classifying the whole many-body eigenstate spectrum of the Hamiltonian. In addition to providing insights into the critical behavior of the MBL transition, these findings also demonstrate the the efficiency of machine learning classification in that it can operate with much less labeled data which are expensive in computation. Thus, when appropriately applied, the SVMs can be more powerful tool for classifying physical data compared to conventional methods, especially in complex physical situations. In the

appendix, we also trained the 3-layer neural networks (NN) machine on the same training sets, and used it in the same way as the SVM to classify the MBL and thermal phases. The phase diagram obtained by neural network machine agrees to that determined by the SVM within the error bars, demonstrating that different machine learning models lead to the consistent classification results in the disordered quantum Ising spin chain.

A unique advantage of the SVM is its interpretability, which indeed allowed us to interpret how the SVM separates the input data belonging to the different phases. Remarkably, we find that the decision function constructed by the SVM is closely related to the generalized IPR in the many-body Hilbert space. The fact that the interpretable machine learning suggests that IPR may have the ability to identity the MBL transition is a physically significant results in that it relates the failure to thermalize to the Anderson type of localization in the many-body Hilbert space. The consistency between the SVM phase diagram and the one obtained from the variance of the entanglement entropy[1] further supports this intriguing possibility. Introducing the participation entropy to describe the many-body IPR, we further explored this connection by directly comparing the entanglement entropy and the participation entropy and found remarkable similarity in their behaviors. Further studies of the interconnection between these two quantities in larger system sizes are however necessary to reach more definitive conclusions.

CHAPTER III

Critical behaviors of the entanglement and participation entropy near the many-body localization transition in a disordered quantum spin chain

3.1 Introduction

In 1958, Anderson proposed that all single particle states of a closed non-interacting quantum system can be localized with sufficient randomness and thus have zero conductivity. Such systems fail to reach thermal equilibrium even after an infinitely long time evolution[41]. About half a century later, Basko, Aleiner, and Altshuler argued that when weak interactions are present, the localization remains[6] in the many-body localization (MBL) phase, which has since then been widely studied theoretically and numerically[1, 3–5, 16, 24–29, 59–61], and observed experimentally in cold atom and trapped ion systems[62–66].

The transition between the MBL phase and the delocalized thermal phase is a dynamical phase transition. As the disorder strength increases, the delocalized system whose long-time behaviour obeys equilibrium thermodynamics turns nonergodic and thus fails to thermalize. Due to the breakdown of ergodicity in

the MBL phase, the transition cannot be described by the conventional quantum statistical-mechanics with averages over many eigenstates. In contrast to ordinary quantum phase transitions that occur in ground states, the MBL transition is a transition in the the many-body eigenstates at finite energy densities. The excited eigenstates that satisfy eigenstate thermalization hypothesis (ETH) are separated from those that fail to satisfy ETH by the MBL transition. A significant implication of ETH is the volume law of the entanglement entropy, whereas it obeys the area law in the MBL phase due to the locality of the interactions, akin to quantum ground states. As a result, the entanglement entropy is widely used as an "order parameter" to study the MBL-ETH phase transition.

Although great progresses have been made, some fundamentals of the MBL-ETH phase transition are still not clear, partly because the MBL transition falls outside equilibrium statistical mechanics. Grover argued [67] that the critical eigenstates are thermal assuming that the MBL-ETH transition is continuous. While some numerical supports for this analysis have been reported[1, 16, 68], other works suggest the behavior at the critical point to be more like that of a localized phase than an ergodic phase[59, 60]. Arguments against the continuous transition assumption have also proposed[17]. Moreover, following the assumption of a continuous MBL transition, a Harris criterion type of bound $\nu \geq 2/d$ for the critical exponent of the divergent length scale has been proposed[69]. While this bound is corroborated by the perturbative renormalization group (RG) studies[70–72], it is violated by essentially *all* current exact diagonalization (ED) and scaling of finite-size systems[1, 16]. Recently, the validity of the Harris criterion for MBL transition has been challenged and an exact result of $\nu = 1$ was derived from the theoretical analysis[73].

Another intriguing and relevant question is wether the MBL ultimately arises through the localization of the many-body states in the configurational Hilbert

space of the entire system L, in analogy to the single-particle Anderson localization in the physical space. Is the entanglement entropy, which is defined on a subsystem $L_A \in L$, and its volume vs area law in the ETH and MBL phases in the thermodynamic limit $L \gg L_A \gg 1$ just one of the many ways for describing the MBL transition? To be specific, consider an exponentially many expansion coefficients of an eigenstate wave function on the thermal side of the transition over some local basis states. Will the number of expansion coefficients be of lower order on the MBL side? If this is the case, there must exist a quantity defined in that local basis space, in analogy to the inverse participation ratio (IPR) that measures the (inverse) of the spatial coverage of the single-particle eigenstates. This quantity can be called as the many-body IPR (mIPR). Does the mIPR exhibit critical behavior near the MBL transition? The mIRP is clearly basis dependent, but is its critical behavior (if any) basis dependent? Is there a naturally specified choice of basis wherein the MBL transition can be described by the mIRP? Several theoretical studies have shown that the behavior of the mIPR or its logarithm termed as the participation entropy and the entanglement entropy share similarities[44, 45], but if the former is a critical quantity in the MBL transition is still under debate.

In this work, we report the progress made on a disordered transverse field Ising chain defined in Section IIA, which is known to display the MBL-ETH transition[1], using an improved ED and a new finite-size scaling analysis that provide several useful insights into these fundamental issues associated with the MBL transition. Specifically, we apply twisted boundary conditions (TBC) that significantly reduce the finite size effect discussed in Section IIB. The ED is carried out on otherwise identical Ising chains where the end spin is rotated by an angle θ around x axis. The relevant quantities are averaged over different twisted angles, disorder realizations, and a small energy density window. We find that this algorithm greatly reduces the finite size effect. We then study the behavior

of the entanglement entropy near the transition in Section III, based on the same two assumptions made by Grover[67]. The first assumption is that the MBL-ETH transition is continuous. This implies, according to Grover's analysis, that the critical entanglement entropy equals to a thermal entropy. The second is that the entanglement entropy S_E^A of the subsystem is a scaling function only of L_A/ξ, with no significant dependence on the total system size L when $L \gg L_A, \xi$. This is reasonable because in the thermodynamic limit, the exact size of the whole system L that acts as a heat bath of the subsystem should not significantly impact the value of S_A. We show that L influences S_E^A only through the dimensionless partition ratio $r = L_A/L$. In the relevant thermodynamic limit where $r \to 0$, we find that S_E^A is strictly thermal at critical point, instead of being subthermal as suggested in Ref[1, 16, 17]. This is consistent with Grover's analysis and therefore corroborates the assumption of the continuity of S_E^A. Following the above analysis, we perform a finite size scaling analysis of S_E^A with L_A/ξ as the scaling variable, whereas L enters through r as corrections to scaling due to irrelevant operators. In this way we find a critical exponent $\nu = 0.94 \pm 0.07$. This value still violates the Harris bound, but agrees well with the result derived in Ref[73]. Finally, in Section IV we perform a finite size scaling analysis of the participation entropy, i.e. the logarithm of the mIPR, defined in a suitable spin configuration space of the domain wall excitations. We find that both the critical point and the critical exponent agree with those obtained from the scaling of entanglement entropy. This result implies that MBL-ETH transition is a localized-delocalized transition in the spin configuration space.

3.2 Model and Methods

3.2.1 Transverse-field disordered quantum Ising chain

The quantum transverse-field Ising chain is known to develop the MBL phase when the disorder strength is strong enough. The Hamiltonian of the system is given by [1]

$$\hat{H} = -\sum_{i=1}^{L-1} J_i \sigma_i^z \sigma_{i+1}^z + J_2 \sum_{i=1}^{L-2} \sigma_i^z \sigma_{i+2}^z + h \sum_{i=1}^{L} \sigma_i^x \tag{3.1}$$

where σ^x and σ^z are Pauli matrices and L is the number of sites in the chain. In Eq. (3.1), the second nearest neighbor coupling J_2 and the transverse external field h are uniform, whereas the nearest neighbor coupling is site-dependent. We use $J_i = J + \delta J_i$, where J is a constant and δJ_i is randomly taken from a uniform distribution $[-\delta J, \delta J]$. Thus δJ measures the disorder strength. For a certain disorder realization, the energy E of the many-body eigenstates of H is bounded within a bandwidth $W = E_{max} - E_{min}$. Consider a disordered ensemble of H, the appropriate dimensionless energy is defined by the energy density $\epsilon = 2(E - E_{min})/W$ relative to the bandwidth, within a small window around ϵ. All quantities computed later are averaged over different disorder configurations in a small energy window around a fixed $\epsilon = 59/60$. We set $J_2 = 0.5h = 0.3J$, where the ground state of the Hamiltonian has ferromagnetic order in z direction due to Z_2 symmetry breaking. As we focus on excited states, in the absence of interactions and randomness, the ferromagnetic order is destroyed at any finite energy density as the excited domain walls are extensive over the whole chain. When there's randomness but no interaction, the domain walls become localized for infinitesimal disorder. With both interaction and randomness, there exists a finite critical disorder strength separating the ETH phase where the domain walls are extensive from the MBL phase where they're localized[1, 3].

3.2.2 Twisted bondary conditions

When applying exact diagonalization (ED) to the Hamiltonian to compute the eigen wavefunctions, we normally use open boundary conditions (OBC), namely, ignore the interactions between boundary spins; or periodic boundary conditions (PBC) which allow the boundary spins to couple in the same way as inner spins; or a mixture of the above. If the system size is large enough, such boundary effect will have negligible impact. However, due to the exponential increase of numerical demand, we can only study systems of very limited sizes. $L = 16$ is the largest system size commonly studied using ED for quantum Ising chains. This limitation leads to strong finite size effects, which become more severe for long-ranged interactions.

To reduce the finite size effects, it is common to use twisted boundary conditions (TBC)[74, 75]. In this paper, we implement TBC by rotating the last spin at the end of the Ising chain by an angle θ around the x axis. This corresponds to a transformation in the Hamiltonian

$$\sigma_L^z \to e^{i\theta S_L^x/\hbar}\sigma_L^z e^{-i\theta S_L^x/\hbar} \tag{3.2}$$

or equivalently,

$$\begin{aligned} \sigma_1^z\sigma_L^z &\to cos(\theta/2)\sigma_1^z\sigma_L^z + sin(\theta/2)\sigma_1^z\sigma_L^y, \\ \sigma_2^z\sigma_L^z &\to cos(\theta/2)\sigma_2^z\sigma_L^z + sin(\theta/2)\sigma_2^z\sigma_L^y, \end{aligned} \tag{3.3}$$

respectively in the nearest and second nearest neighbor coupling terms. The coupling term between σ_L^x and the transverse field are kept unchanged. When $\theta = 0$, the last spin is along the z axis, which is simply PBC; and when $\theta = 2\pi$, it correspond to anti-PBC. We vary θ from 0 to 2π. For each θ value, we use a number of

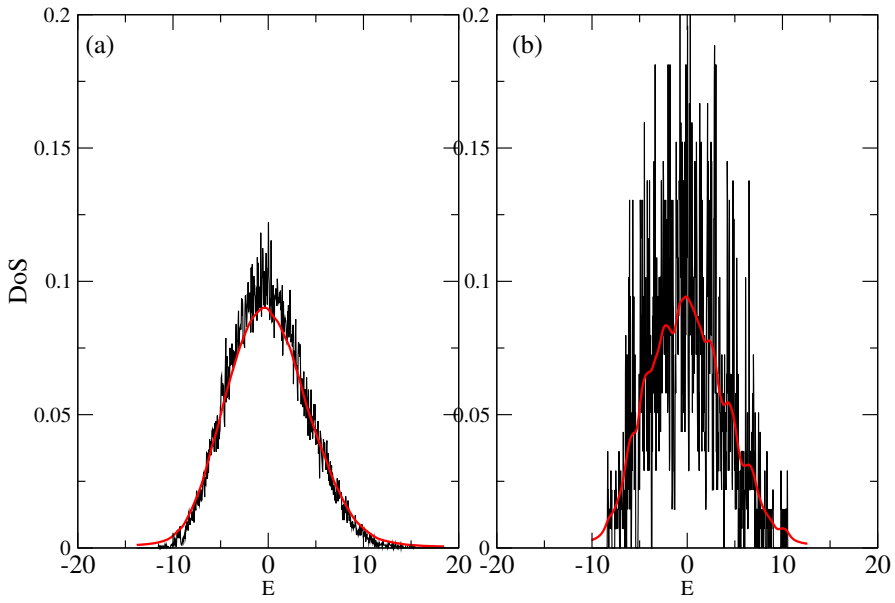

Figure 3.1: Averaged DoS for the disordered spin chain with $L = 8$ computed using TBC (a) and PBC (b). Red lines are obtained from spectral function and black lines from counting the number of eigenstates in a small binned energy window. They are obtained by averaging over the same number of ED results and presented under the same parameter settings.

disorder configurations to generate eigen-wavefunctions. All quantities computed from the wavefunctions will be averaged over both different θ and different disorder configurations within a small energy window around an energy density, which will be discussed in detail below.

Fig.3.1 shows that the application of TBC greatly improves the smoothness of the density of states (DoS). Fig.3.1(a) and Fig.3.1(b) are the DoS obtained using TBC and PBC respectively, with both of them averaged over the same number of ED results. The red line denotes the DoS computed from the spectral function and the black line is obtained by counting the number of eigen-energies in a binned energy window. The values of the infinitesimal real positive number in the spectral function and the width of the energy window are chosen to be the same in Fig.3.1(a) and Fig.3.1(b).

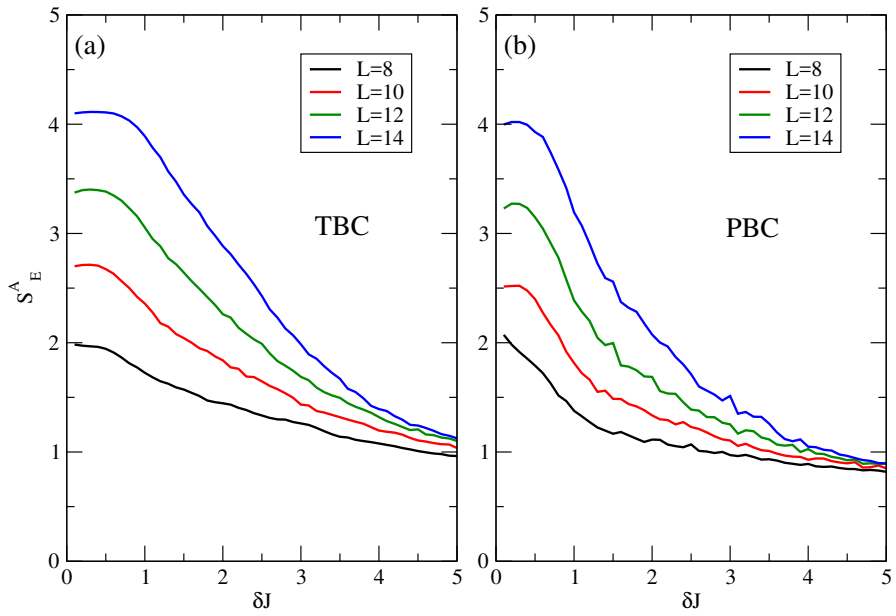

Figure 3.2: Entanglement entropy as a function of disorder strength computed for spin chains of lengths $L = 8, 10, 12, 14$ under TBC (a) and PBC (b). The average is taken over the same number of disorder configurations.

In Fig.3.2 and Fig.3.3, we show that averaging over TBC compared to using pure PBC also reduces the fluctuations of the quantities computed from the wave functions, thus benefits the scaling analysis to be discussed later. We take entanglement entropy and participation entropy (which will be detailed in the following sections) as examples. Fig.3.2(a) shows that the entanglement entropy calculated with TBC has much less fluctuations as a function of the disorder strength, compared to that calculated using PBC in Fig.3.2(b). Fig.3.3 illustrates the same effect for the participation entropy.

From the above results, we find that using TBC can greatly mitigate the boundary effects in finite size systems as it removes the bias of fixing a specific boundary condition and improves the quality of the numerical data by reducing fluctuations of averaged quantities. In the rest of the paper, we will use TBC exclusively to generate all required data.

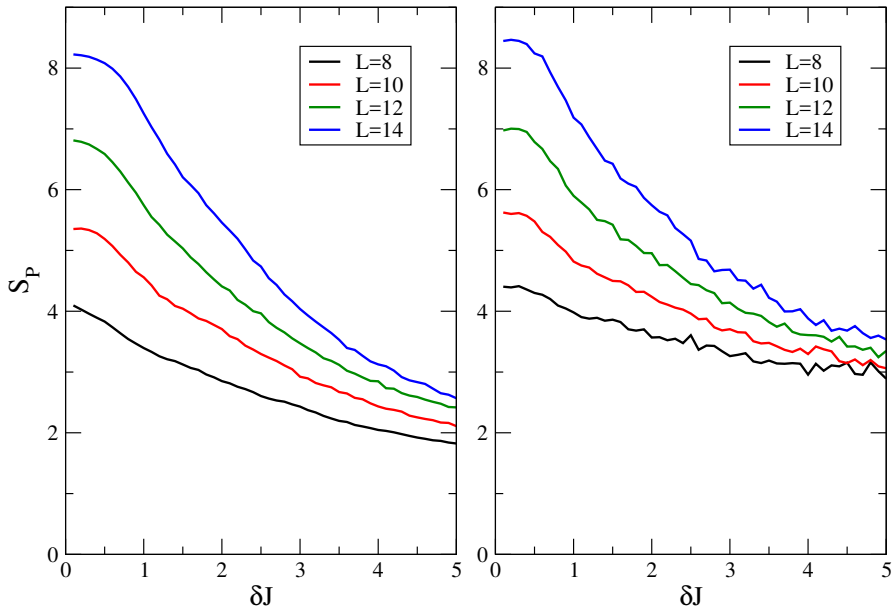

Figure 3.3: Participation entropy as a function of disorder strength computed for spin chains of lengths $L = 8, 10, 12, 14$ under TBC (a) and PBC (b). The average is taken over the same number of disorder configurations.

3.3 Entanglement Etropy

3.3.1 Qualitative behavior

We now study the qualitative behavior of the entanglement entropy $S_E^A(L_A, L, \delta J)$ of a subsystem A of length L_A, where L is the length of the whole spin chain and δJ is the disorder strength. It is given by

$$S_E^A = -\mathrm{Tr}_A \rho_A \ln \rho_A, \tag{3.4}$$

where

$$\rho_A = \mathrm{Tr}_{A^c} |\Psi_n\rangle\langle\Psi_n| \tag{3.5}$$

is the reduced density matrix with the trace Tr_{A^c} running over the complement set of A, i.e. $L - L_A$. $|\Psi_n\rangle$ is the n^{th} eigenstate with energy E_n obtained by exact diagonalization of the Hamiltonian in Eq. (3.1). In this work we focus

on eigenstates with energy densities falling into a small energy window around $\epsilon = 59/60$. It is already clear that in the ETH phase, S_E^A obeys the volume law, namely it is proportional to the *subsystem size*. Therefore in our one-dimensional system, it is linear in L_A. On the other hand, in the MBL phase, S_E^A should obey the area law, being proportional to the boundary of the subsystem, i.e. it should be a constant in the one-dimensional case. In previous works on the spin chains, it is common to study the half-chain entanglement entropy and its scaling behavior with respect to the whole system size L[1, 16]. Doing so does not clearly addressed if the entanglement entropy scales with L or L_A. Although in this case scaling with L is equivalent to scaling with L_A, discussing the scaling behavior with respect to the total system size L may be confusing. There are two reasons. First of all, by definition, S_E^A should only scale with L_A, since when L is large enough to satisfy $L \gg L_A$ and $L \gg \xi$ where $\xi \sim |\delta J - \delta J_c|^{-\nu}$ is the correlation length, S_E^A should be independent of the value of L. This can be understood since in the thermodynamic limit $L \gg L_A \gg 1$ because if the subsystem in the ETH phase can thermalize and the rest of the system is able to act as an infinite heat bath, the exact size of the heat bath should not matter. If instead, the subsystem is in the MBL phase, the area law implies that the one-dimensional systems will be independent of either L_A or L. Thus the only dimensionless scaling variable should be L_A/ξ. Second, requiring $L_A = L/2$ is far away can never really approach the $L_A \ll L$ limit. As a result, the finite size effect may greatly impact the behavior of S_E^A and thus systems may be more difficult to be fully thermalized and thus more prong to subthermal behavior of S_E^A. In the ED studied of small to moderate finite size systems, the partition ratio $r = L_A/L$ can enter through corrections to scaling which must be taken into account.

We further illustrate the above statements by calculating the entanglement entropy at fixed $L_A = 2$ and explore how S_E^A varies with L for different disorder

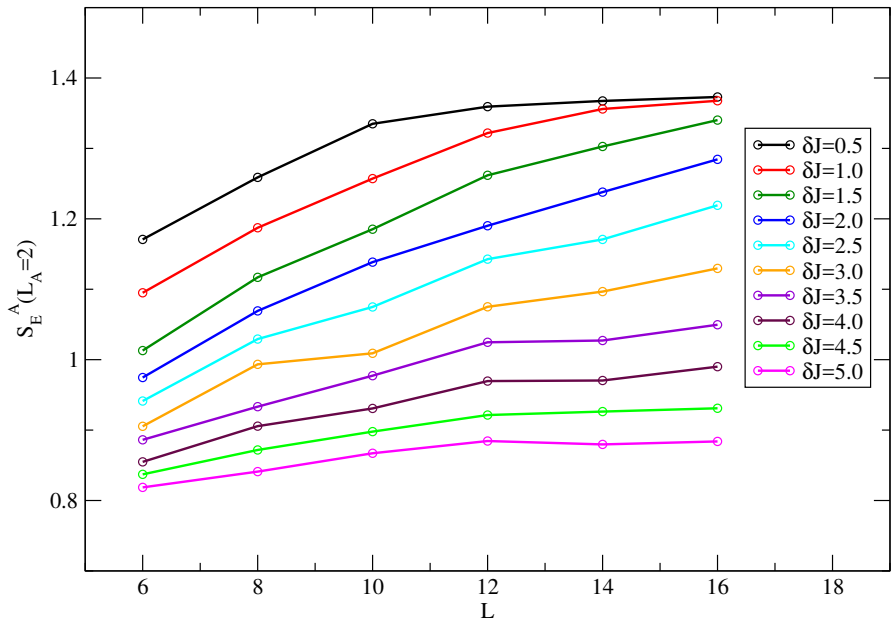

Figure 3.4: The entanglement entropy S_E^A as a function of L at fixed subsystem size $L_A = 2$ for different δJ. When L is large, S_E^A is approximately independent of L for both small δJ smaller than 1.0) and large δ (larger than 4.5). In the region $1.5 \leq \delta J \leq 4.5$, S_E^A shows dependence on L even when for the largest $L = 16$.

strength δJ. The results of the ED are shown in Fig. 3.4. It can be seen that when $L > 12$, for very weak and very strong disorder strengths, i.e. $\delta J \leq 1.0$ and $\delta J \geq 4.5$ respectively, S_E^A shows approximately no dependence on L, implying that $L \gg L_A$ and $L \gg \xi$ are both satisfied. On the other hand, if $L \leq 12$ in the region of disorder strength above, or for $1.0 < \delta J < 4.5$ and all $L \leq 16$, S_E^A appears to increase with L. This latter behavior is likely caused by the violation of $L \gg \xi$, as ξ becomes larger when δJ gets closer to its critical value. First, we emphasize that this increase is a finite size correction rather than the volume law in L. Because S_E^A is bounded in this case by $\ln 2 L_A$, the dependence of S_E^A on L must vanish for large enough L. Second, the finite size correction in the regime where L is comparable to ξ reduces the value of S_E^A, which should be accounted for. To achieve this goal, we introduce the partition ratio $r = L_A/L$. When $r \ll 1$, meaning that when $L \gg L_A$, this correction is negligible. When r is finite, it adds a correction to the scaling of S_E^A in the regime around the critical point with ξ, in the same spirit as correction to scaling due to irrelevant operators. By doing so we express all L dependence on S_E^A through r, thus the entanglement entropy $S_E^A(L_A, L, \delta J) = S_E^A(L_A, r, \delta J)$. The quantitative study of its scaling behavior will be discussed in the next section. Next, we examine the dependence of S_E^A on L_A and demonstrate why the bipartition entanglement entropy, namely $S_E^A(r = 0.5)$, may not be a good choice for studying the scaling behavior of S_E^A. In Fig. 3.5 we show how S_E^A varies with r for different fixed L at weak, moderate and strong disorder strengths. Fig. 3.5a-c show that at small or moderate δJ, S_E^A shows linear dependence in r for fixed L, i.e. linear in L_A, when r is small, implying a volume law of S_E^A in L_A. As r increases and approaches 0.5, S_E^A's linear dependence on L_A becomes invalid. The condition that r must be small for the volume law to hold is consistent with the thermodynamic limit requirement $L \gg L_A$, i.e. when the rest of the system can act as an infinite reservoir for the subsystem. Since

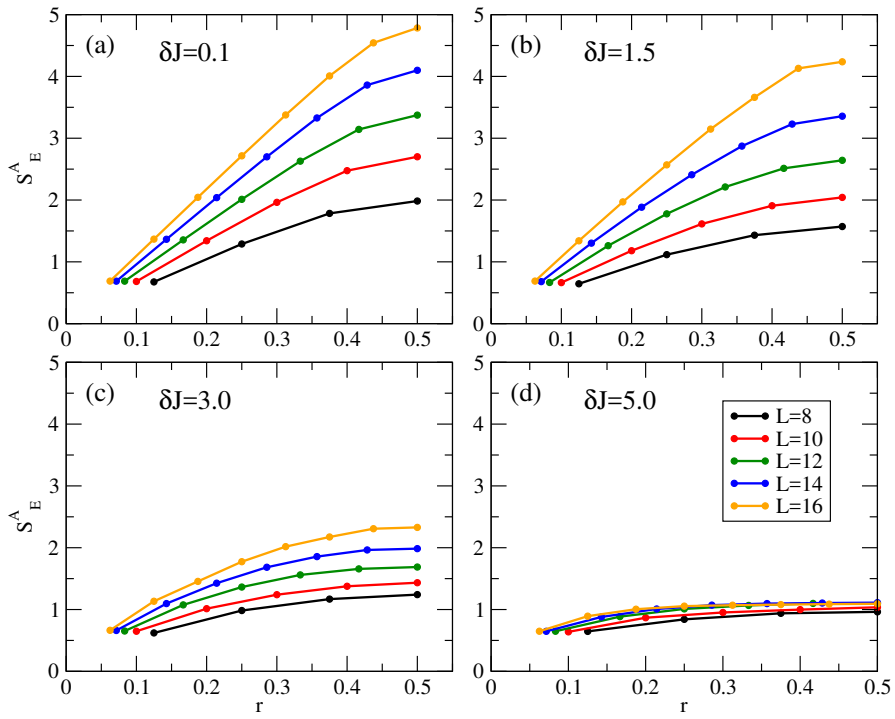

Figure 3.5: Entanglement entropy as a function of the partition ratio $r = L_A/L$ at a fixed L for 5 different system sizes $L = 8, 10, 12, 14, 16$. Each panel corresponds to a fixed disorder strength: $\delta J = 0.1$ (a), 1.5 (b), 3.0 (c), and 5.0 (d).

S_E^A seems not to be fully thermalized at $r = 0.5$ due to the small whole system size relative to the subsystem size and do not obey volume law even at very weak disorder strength as seen in Fig. 3.5a, the bipartition entanglement entropy may not be an appropriate choice for studying the scaling behavior. Fig. 3.5d shows that when δJ is large ($\delta J = 5.0$), the dependence of S_E^A on either L or L_A is rather weak, which is consistent with the area law behavior.

The above observations are qualitatively consistent with the analysis of Grover [67] and support the notion that the MBL transition is continuous in the thermodynamic limit. We note that in ref[17], it was argued that in the quantum critical regime, the entanglement entropy obeys the area law, which challenges the assumption that the MBL-ETH transition is continuous. We also observed similar plateaus in the S_E^A versus L plot at a small fixed L_A in Fig. 3.4. However, since

the volume or area law is defined with respect to L_A instead of L, we refrain from regarding the plateau as an indication of the area law in quantum critical regime. Moreover, it can be seen in Fig. 3.5c where $\delta J = 3.0$, S_E^A for $L = 8, 10$ corresponds to the plateau region in Fig. 3.4, the entanglement entropy as a function of L_A appears not to behave very differently from those at $L = 12, 14, 16$. As a result, we believe the continuity assumption is still valid, and based on which we proceed to build our scaling theory.

3.3.2 Finite-size scaling for S_A

The finite-size scaling for the entanglement entropy in the MBL phase transition has been widely studied[1, 16, 17, 67, 76]. Based on the above observation and discussion, we need to construct a scaling theory explicitly in the subsystem sized L_A, such that when $\delta J < \delta J_c$, S_E^A obeys the volume law in L_A therefore S_E^A/L_A is a finite constant; when $\delta J > \delta J_c$, S_E^A is a constant independent of L_A thus $S_E^A/L_A \to 0$ when $L_A \gg 1$, following area law in L_A; at the critical point $\delta J = \delta J_c$, S_E^A is continuous and thus maintain a critical volume law. To this end, we specify the system and subsystem sizes using (L_A, r) and all correction to the scaling of the entanglement entropy caused by a finite L is expressed in terms of the noncritical dependence on r. The entanglement entropy S_E^A thus depends on three variables as $S_E^A(L_A, r, \delta J)$, or equivalently, $S_E^A(L_A, r, \xi(\delta J))$ where ξ is the correlation length. The finite-size scaling form can therefore be written according to

$$S_E^A(L_A, r, \delta J) = L_A f(L_A|\delta J - \delta J_c|^\nu, r). \tag{3.6}$$

The scaling function $f(x, r)$ can be expanded in the vicinity of the critical point,

$$S_E^A/L_A = f_c(r)(1 + g_r(L_A|\delta J - \delta J_c|^\nu)), \tag{3.7}$$

where $f_c(r) = f(0, r)$ is the value of S_E^A/L_A at the critical point δJ_c. The function $g_r(x)$ can be expanded as a polynomial in x with $g_r(0) = 0$; its expansion coefficients have in general r dependence.

The functional form in Eq. (3.7) suggests that at the critical point, for all combinations of L and L_A, S_E^A/L_A should collapse to the same function $f_c(r)$. In this way we can determine the critical disorder strength by choosing δJ_c to be the point where all S_E^A/L_A (for example, all data points in of Fig.3.5 divided by their corresponding subsystem size L_A) collapse the best. The resulting $f_c(r)$ describes the critical volume law amplitude at the transition for any given r. In Fig. 3.6, S_E^A/L_A is shown as a function of r on scatter plots for all L at different disorder strength ΔJ. We find that the best data collapse arises at $\delta J_c = 3.2 \pm 0.1$ corresponding to Fig. 3.6c, which gives the critical disorder strength for the transition.

The functional curve of the collapsed data in Fig. 3.6c provides the critical amplitude function $f_c(r)$. Fitting the data according to

$$f_c(r) = \alpha + \beta r + \gamma r^2 \tag{3.8}$$

we obtain $\alpha = 0.72 \pm 0.03$, $\beta = -1.31 \pm 0.22$, and $\gamma = 0.84 \pm 0.62$ with the mean square root fitting error equal to $9.98 * 10^{-3}$. This allows us to extrapolate the critical amplitude to the thermodynamic limit $1 \ll L_A \ll L$ by taking the limit, $\lim_{r\to 0} f_c(r) = a = 0.72 \pm 0.03$ for the prefactor of the volume-law at critical point. Remarkably, this result agrees within the error bar with the thermal entropy which is $ln2L_A$ in the high temperature limit, where the prefactor is $\ln 2 \approx 0.693$. This serves as a self-consistency check and confirms that S_E^A is continuous and obeying the volume-law in L_A at transition. Subthermal behavior of S_E^A near the critical point has been observed previously[1, 16, 77, 78], but these works do not focus

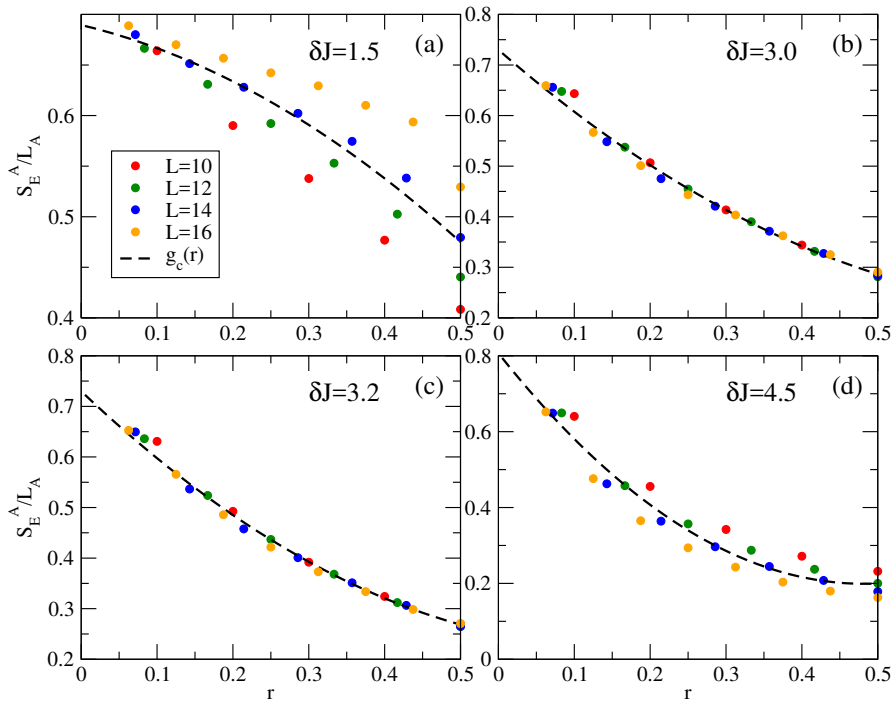

Figure 3.6: Scatter plots of S_A/L_A as a function of $r = L_A/L$ for $\delta J = 1.5$ (a), 3.0 (b), 3.2 (c), and 4.5 (d). There are 26 data points obtained for $L = 10, 12, 14, 16$ and $L_A = 1, 2, \cdots, L/2$. The least squared fits are shown by dashed lines.

on the limit $L_A \ll L$. As can be seen from Fig. 3.6c, $f_c(r) < \ln 2$ for nonzero r, strongly suggesting that the subthermal behavior near the transition is due to a failure to satisfy the $L \gg L_A$ limit that is most severe at the half-chain partition $r = 0.5$. In this case, the ability of the rest of the system to act as a heat bath is impaired, making the subsystem not fully thermalized.

We now turn to the finite size scaling analysis of the entanglement entropy S_E^A. Instead of scaling S_E^A/L_A in Eq. (3.7), we scale the quantity $y = S_E^A/(L_A f_c(r))$, where $f_c(r)$ has already been obtained by the above procedure and given in Eq. (3.8). Thus $y = 1 + g_r(x)$ where $x = L_A|\delta J - \delta J_c|^\nu$. Since the major dependence of on r has been removed in y, we assume the remaining r dependence in the function $g_r(x)$ is small and negligible in the neighborhood of the critical point, provided that r is small and not very close to 0.5. This assumption can be

further discussed below.

This algorithm allows us to choose the largest system size studied namely $L = 16$ and $L_A = 1, 2, 3, 4, 5, 6, 7$ to complete the finite size scaling analysis. The obtained scaling plot is shown in Fig. 3.7. The good quality of the data collapse allows us to determine the critical point $\delta J_c = 3.19 \pm 0.03$, which is consistent with the condition under which $f_c(r)$ was obtained. The critical correlation length exponent is $\nu = 0.94 \pm 0.067$. Note that the data for $L_A = 8$ are not included, because it corresponds to $r = 0.5$ which is too large for the coefficients in $g_r(x)$ to be treated as independent of r. We checked that if $L_A = 8$ were included in the scaling plot, the quality of the data collapse becomes noticeably poorer. In fact, our theory implied that the value of L should not matter if only we divide S_E^A/L_A by the corresponding $f_c(r)$. To verify this point, we repeated the finite size scaling analysis for $L = 14$ and $L_A = 1, 2, 3, 4, 5, 6$ and obtained a similar quality of data collapse with the critical point $\delta J_c = 3.14 \pm 0.06$ and the correlation length exponent $\nu = 0.99 \pm 0.09$, which are very close to obtained when using $L = 16$. The errors in our analysis are obtained following the approach presented in the supplementary material of reference[16]. The good quality of the data collapse and the similar results obtained for different system sizes justify that the r dependence in $g_r(x)$ near the critical point ($x = 0$) is negligible when r is away from 0.5.

3.4 Participation Entropy

Finally, we investigate if the MBL transition describable by the entanglement entropy of a subsystem L_A in a system L in the limit $1 \ll L_A \ll L$ can be characterized by the inverse participation ratio (IPR) or its associated participation entropy in the spin configuration space of the entire system L. In other words, is the MBL related to the localization of the eigenstates in the spin configuration space of the entire system L due to strong disorder and correlation?[44, 45, 79, 80] The con-

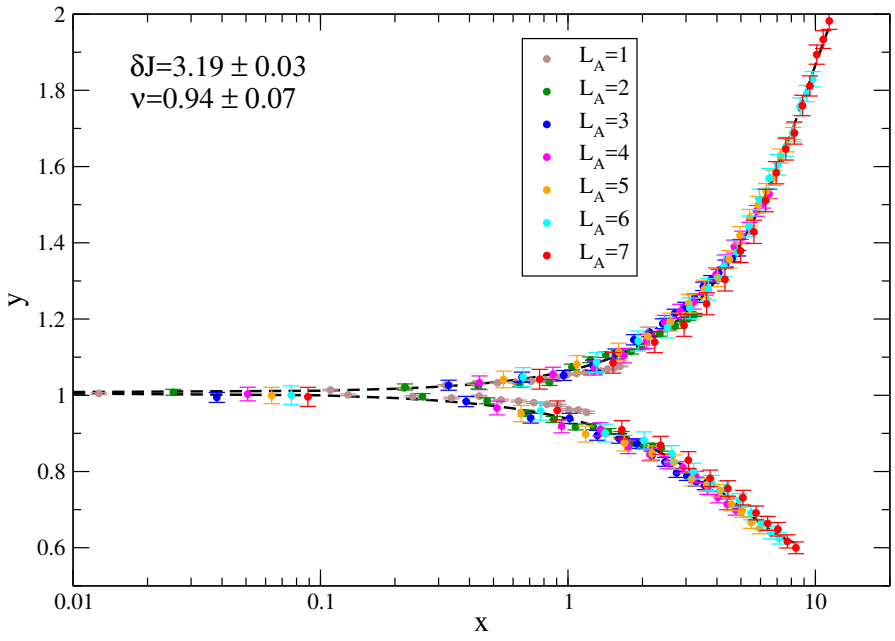

Figure 3.7: Finite size scaling plot of $y = S_E^A/(f_c(r)L_A)$ versus x at $L = 16$, where $f_c(r)$ is determined in Fig 3.6c and $x = L_A|\delta J - \delta J_c|^\nu$. The scaling function is expanded as a polynomial in x.

nections between Hilbert space localization and energy level statistics have indeed been explored and demonstrated by experimental measurements recently[48]. The IPR, which is widely used to study the Anderson localization transition[57, 81], can be generalized to the many-body probelm[58], defined for the n^{th} eigenstate $|\Psi_n\rangle$ as,

$$I_q(n) = \sum_i |\langle\{\sigma_i^z\}|\Psi_n\rangle|^{2q}, \quad q = 2, 3, 4 \cdots, \tag{3.9}$$

where $\{\sigma_i^z\}$ is chosen as the basis for the spin configuration for the model parameters used. This choice of basis will be discussed below. The associated participation entropy is given by

$$S_P^q(n) = \frac{1}{1-q} \ln I_q(n). \tag{3.10}$$

We consider $q = 2$ and ignore the superscript q for simplicity in the rest of the discussion.

Several theoretical studies have shown that the behavior of S_P and the entan-

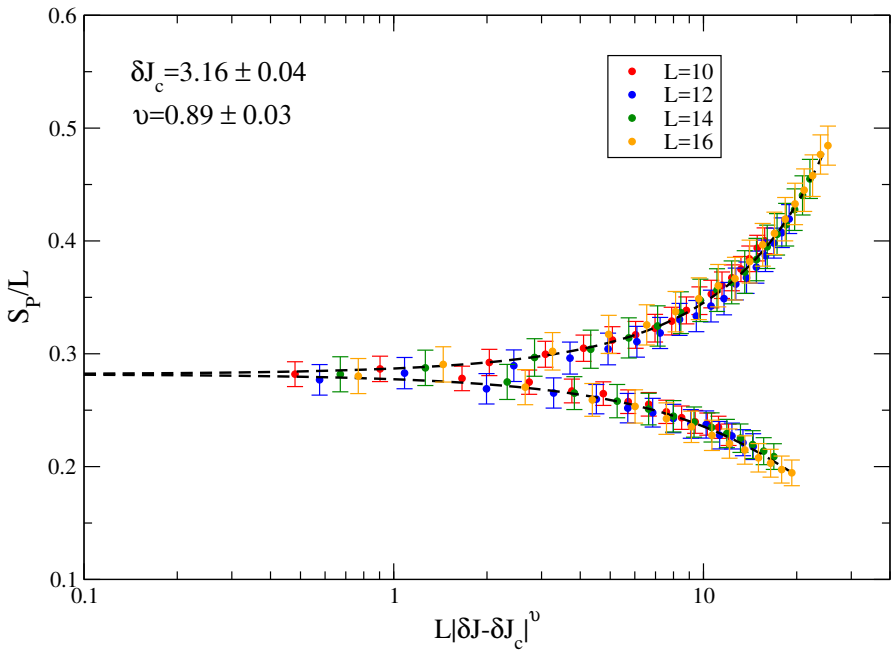

Figure 3.8: Finite size scaling plot of the participation entropy S_P versus $L|\delta J - \delta J_c'|^{\nu'}$. The critical point determined by the data collapse is $\delta J_c' = 3.16 \pm 0.04$ and the critical exponent $\nu' = 0.89 \pm 0.03$.

glement entropy share certain similarities[44, 45, 79] and are directly related to each other in the single-particle case[46], while others offer opposite arguments[16, 47]. Using the high quality S_P data obtained by ED under the TBC, we performed a finite size scaling analysis of S_P obtained on finite-size systems with $L = 10, 11, 12, 13, 14, 16$ at different disorder strengths. The scaling plot is shown in Fig.3.8, which gives the critical disorder strength $\delta J_c' = 3.16 \pm 0.04$ and the critical correlation length exponent $\nu' = 0.89 \pm 0.03$. Comparing the results to those obtained for the entanglement entropyS_A, we find that $(\delta J_c, \nu)$ and $(\delta J_c', \nu')$ agree within the numerical uncertainty, suggesting that the critical behaviors in the partition entropy S_P and the entanglement entropy S_E^A may describe the same phase transition, and therefore MBL is possibly a localization in the spin configuration space.

Note that the definition of the many-body IPR in Eq. (3.9) requires choosing

a basis and the obtained results can be basis dependent. It is thus important to discuss if there exists a physically suitable basis for studying MBL using the participation entropy. Here we chose the spin configuration basis along the z direction in Eq. (3.9) and demonstrated that the participation entropy S_p in Eq. (3.10) displays a MBL transition consistent with that of the basis-independent entanglement entropy S_E^A for the spin-chain. The reason for this basis choice is that the ground state of the system has ferromagnetic order of the spins along the z direction at $h = 0.6J$. Thus the excited states are domain wall excitations that flip the spin along the z direction[1, 3]. In the ETH phase the domain wall excitations are extensive, whereas in the MBL phase they are localized. There exists an occupation number representation for the domain walls in terms of the local spin configuration in the z direction. For example, in our model system with Z_2 symmetry under $\hat{P} = \prod_{i=1}^{L} \sigma_i^x$ and OBC, an unoccupied domain wall state $|0\rangle$ is given by $(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle)/\sqrt{2}$, while an occupied domain wall state $|1\rangle$ corresponds to $(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)/\sqrt{2})$. We believe that the suitable basis to define the many-body IPR is the quasiparticle occupation number basis. Thus, for the parameter space leading to the ferromagnetic ground state and domain wall excitation, the spin configuration in the z direction is the suitable choice of basis with which we computed the participation entropy. This implies that when $h \gg J$ where the domain wall excitations turn to flip spins in the x direction, the appropriate choice of basis would be the spin configuration in the x direction. While more detailed studies are necessary, we have tested this at $h = 4J$ and found that the participation entropy S_P defined in the spin configurations along the x direction shows the scaling behavior, while that along the z direction does not. Finally, any two choices of basis connected by transformations that commute with the Hamiltonian will give the same behavior of the m-IPR. In the present case, identical results are obtained if the basis is chosen to be the spin configuration along the $-z$ direction. In the

more general case where the Hamiltonian has $SU(2)$ symmetry, the basis choices of spin directions along x, y or z will produce the same results.

3.5 Summary

We introduced a finite size scaling theory for the critical behavior of the entanglement entropy near the MBL transition. We emphasized that it is the subsystem size L_A, not the total system size L, that cuts off the critical singularity of the MBL transition, such that the only scaling variable is L_A/ξ with ξ being the divergent correlation length. The total system size L enters only as a correction to scaling through the introduction of the partition ratio $r = L_A/L$ that characterizes the ability of the rest of system to act as an infinite heat bath to the subsystem. While r vanishes in the thermodynamic limit $L \gg L_A$, a large r close to its maximum $r_m = 0.5$ pronounces the inability to fully thermalize, leading to subthermal behaviors of the entanglement entropy near the critical point. In the thermodynamic limit, the scaling function produces the volume and area laws of the entanglement entropy in L_A but not in L in the ETH and MBL phases, respectively. We applied this scaling theory to the MBL transition in the disordered transverse field Ising chain. The finite size scaling analysis of the entanglement entropy, obtained using exact diagonalization and twisted boundary conditions to reduce the boundary effects, supports that the MBL-ETH transition is continuous and the entanglement entropy is strictly thermal at the critical point. The correlation length exponent is obtained to be $\nu = 0.94 \pm 0.07$. This value is below the lower bound set by the Harris criteria, as are several other exact diagonalizaiton results. It is, however, close to $\nu = 1$ derived by Monthus[73] who argued that the critical exponent for the MBL transition need not satisfy the Harris bound since the quantum many-body state under study experiences $2^L - 1$ random energies, which is much more than L^d assumed in the derivation of the Harris bound. It remains to be explored if

the Harris bound is violated due to the limited system size, or because it does not apply to the MBL phase transition studied. We also argued that the many-body IPR and the participation entropy defined in the quasiparticle occupation number basis of domain wall excitations in the entire system L can be used to describe the MBL transition as a localization transition of the many-body eigenstates in the spin configuration space. We find that the finite size scaling analysis of the participation entropy results in a critical point and a correlation length exponent very close to those obtained from the entanglement entropy, suggesting they share similar critical behaviors near the MBL transition. In addition to providing new insights to the subject of MBL, this scaling theory should be applicable to other models of MBL.

CHAPTER IV

Finite Temperature Gutzwiller Approximation of Hubbard Model and the Equation of States

4.1 Introduction

The Hubbard model works as a paradigm for studying electronic systems with strong but short-range Coulomb interactions. Despite its simple form, it strikingly captures most important aspects of the system on one hand, on the other hand, it lacks anaylitical solution except in one-dimension. The Gutzwiller projection methond, sometimes in combination with variational Monte Carlo approach, is widely used to treat the Hubbard model at zero temperature. The reasons for its popularity lie in two aspects, firstly it characterize the fact that the bandwidth is renormalized (narrowed) by the interaction, which other traditional mean field methods fail to account. Secondly, it describes the metal-insulator transition driven by strong Coulomb interaction. The most successful application of Gutzwiller projection is in the study of ground state, one can variationally search for the ground state by devising it as $|G\rangle = P|G_0\rangle$, where P is the Gutzwiller projector and $|G_0\rangle$ is a suitable non-interacting state. Equivalently, this can be regarded as finding a renormalized noninteracting Hamiltonian $\hat{H}_{GW}$ to approximate the interacting Hamiltonian. It's then natural to ask if this projection

method can be extended to finite temperature. While in contrast to zero temperature, this requires to treat an ensemble instead of one pure state, thus we need to generalize the projection from the wavefunction level to the density matrix level.

In this work we construct a density matrix operator to approximate the real interacting density matrix, since the real one gives the lowest energy, our constructed one can at least provide an upper bound of the free energy. The constructed density matrix can be anything if only it's positive definite and its trace is unity, we take such an ansarts, where P is still of Gutzwiller type. We get the exact form of internal energy and a lower bound of entropy, thus the upper bound of free energy. The entropy contribution is from the projector, which is the incoherent part that cannot be characterized by quasiparticles. We apply this method to half-filled Hubbard model to study the metal-insulator transition with and without magnetism.

When assuming no magnetic order, we find that there exists a critical temperature T_c, below which there exists two repulsion strength U_{c1} and U_{c2} such that below U_{c1} the system is in phase I; above U_{c2} the system is in phase II; between U_{c1} and U_{c2} the two phases coexist. We'll discuss these 2 phase in detail later. When the temperature is above T_c, the two phases are indistinguishable. We plot the phase diagram and compare our result to that of dynamical mean field theory. We find some qualitative agreements as well as some new physics which will be detailed in the following subsections.

When considering magnetism, we find that the Neel temperature is increasing with repulsion U when U is small, which agrees with the result of Hartree Fock mean field theory. When U is large, T_{Neel} decreases with $1/U$ which agrees with the result of Heisenberg model. More details will be shown in later chapters.

4.2 Extension of Gutzwiller projection to finite-temperature

4.2.1 Internal energy

We know for the interacting system at finite temperature T, the free energy functional is:

$$F(\rho) = (H_{int}\rho) + kT(\rho \ln \rho) \tag{4.1}$$

where ρ is positive definite, satisfying $\rho = 1$.

Obviously, the first part to the $\cdot lhs\cdot$ of equation (A.1) is internal energy, and the second part is $-TS$, where $S = -k\rho \ln \rho$ is the entropy. The exact form of the density matrix is denoted as ρ_{min}, expressed as:

$$\rho_{min} = \frac{e^{-\beta \widehat{H}_{int}}}{Tr(e^{-\beta \widehat{H}_{int}})} \tag{4.2}$$

Therefore the exact free energy is:

$$F_{exact} = F(\rho_{min}) \leq F(\rho), \qquad \forall \rho > 0 \quad with \quad Tr\rho = 1 \tag{4.3}$$

If we choose a variational desity matrix ρ satisfying above condition, by varying ρ we can find an upper bound of free energy.

If we assume the variational ansatz is:

$$\rho = \frac{\widehat{P} e^{-\beta(\widehat{H}_0 - \mu \widehat{N})} \widehat{P}}{Tr(\widehat{P} e^{-\beta(\widehat{H}_0 - \mu \widehat{N})} \widehat{P})} \tag{4.4}$$

where $\widehat{P}$ is a projection operator specified in the grand canonical ensemble and $\widehat{H}_0$ is a non-interacting Hamiltonian. Define the non-interacting average governed by $\widehat{H}_0$ at finite temperature to be $\langle \widehat{O} \rangle_T^0 := \frac{Tr(\widehat{O} e^{-\beta(\widehat{H}_0 - \mu \widehat{N})})}{Z^0}$, where $Z^0 := Tr(e^{-\beta(\widehat{H}_0 - \mu \widehat{N})})$. then $Z := Tr(\widehat{P} e^{-\beta(\widehat{H}_0 - \mu \widehat{N})} \widehat{P}) = \frac{Tr(\widehat{P}^2 e^{-\beta \widehat{H}_0})}{Z^0} \cdot Z^0 = Z^0 \cdot \langle \widehat{P}^2 \rangle_T^0$.

Thus equation (A.2) becomes:

$$\rho = \frac{\widehat{P} e^{-\beta(\widehat{H}_0 - \mu\widehat{N})} \widehat{P}}{Z^0 \cdot \langle \widehat{P}^2 \rangle_T^0} \tag{4.5}$$

If we choose the projection operator to be Gutzwiller type:

$$\widehat{P} = \prod_i (\widehat{E}_i + y_{i\sigma}\widehat{Q}_{i\sigma} + g_i y_{i\sigma} y_{i\bar{\sigma}} \widehat{D}_i) \tag{4.6}$$

where

$$\widehat{E}_i = (1 - n_{i\uparrow})(1 - n_{i\downarrow}) \tag{4.7}$$

$$\widehat{Q}_{i\sigma} = n_{i\sigma}(1 - n_{i\bar{\sigma}}) \tag{4.8}$$

$$\widehat{D}_i = n_{i\uparrow} n_{i\downarrow} \tag{4.9}$$

Then $\langle \widehat{P}^2 \rangle_T^0$ can be factorized as what we did in ground state Gutzwiller projection, the only difference is here we use finite-temperature Wick's theorem. So we get $\langle \widehat{P}^2 \rangle_T^0 = \prod_i z_i(T)$, with $z_i(T) = e_i^0(T) + y_{i\sigma}^2 q_{i\sigma}^0(T) + (g_i y_{i\uparrow} y_{i\downarrow})^2 d_i^0(T)$, where $e_i^0 = \langle \widehat{E}_i \rangle_T^0$, $q_{i\sigma}^0 = \langle \widehat{Q}_{i\sigma} \rangle_T^0$, $d_i^0(T) = \langle \widehat{D}_i \rangle_T^0$.

Thus the corresponding quantities after projection are:

$$e_i(T) = (E_i \rho) = e_i^0(T)/z_i(T) \tag{4.10}$$

$$q_{i\sigma}(T) = (Q_{i\sigma}\rho) = y_{i\sigma}^2 q_{i\sigma}^0(T)/z_i(T) \tag{4.11}$$

$$d_i(T) = (D_i \rho) = (g_i y_{i\uparrow} y_{i\downarrow})^2 d_i^0(T)/z_i(T) \tag{4.12}$$

Now we can evaluate $(H_{int}\rho)$.

$$
\begin{aligned}
(H_{int}\rho) &= (H_{kinetic}\rho) + U(D\rho) \\
&= \sum_{<ij>\sigma} \frac{(\widehat{P}e^{-\beta(\widehat{H}_0-\mu\widehat{N})}\widehat{P}C^{\dagger}_{i\sigma}C_{j\sigma})}{Z^0\cdot\langle\widehat{P}^2\rangle^0_T} + U\sum_i (D_i\rho)
\end{aligned} \tag{4.13}
$$

Firstly we should find a renormalized non-interacting Hamiltonian to take place of the $\cdot lhs\cdot$ of equation (4.13). Ignoring inter-site Wick contraction we can get:

$$
\begin{aligned}
&\frac{(\widehat{P}e^{-\beta(\widehat{H}_0-\mu\widehat{N})}\widehat{P}C^{\dagger}_{i\sigma}C_{j\sigma})}{Z^0\cdot\langle\widehat{P}^2\rangle^0_T} \\
&= \frac{(\widehat{P}C^{\dagger}_{i\sigma}C_{j\sigma}\widehat{P}e^{-\beta(\widehat{H}_0-\mu\widehat{N})})}{Z^0\cdot\langle\widehat{P}^2\rangle^0_T} \\
&= \frac{(\widehat{P}_iC^{\dagger}_{i\sigma}\widehat{P}_i\widehat{P}_jC_{j\sigma}\widehat{P}_je^{-\beta(\widehat{H}_0-\mu\widehat{N})})\prod_{l\neq i,j}z_l(T)}{Z^0\cdot\prod_l z_l(T)} \\
&= \frac{\langle\widehat{P}_iC^{\dagger}_{i\sigma}\widehat{P}_i\widehat{P}_jC_{j\sigma}\widehat{P}_j\rangle^0(T)}{z_iz_j} \\
&= \frac{\langle\widehat{\Gamma}_{i\sigma}\rangle^0_T}{z_i}\frac{\langle\widehat{\Gamma}_{j\sigma}\rangle^0_T}{z_j}\langle C^{\dagger}_{i\sigma}C_{j\sigma}\rangle^0_T \\
&= \langle g^t_{i\sigma}(T)g^t_{j\sigma}(T)C^{\dagger}_{i\sigma}C_{j\sigma}\rangle^0_T,
\end{aligned} \tag{4.14}
$$

where $\widehat{P}_iC^{\dagger}_{i\sigma}\widehat{P}_i = y_{i\sigma}\widehat{Q}_{i\sigma}C^{\dagger}_{i\sigma}\widehat{E}_i + y^2_{i\bar{\sigma}}y_{i\sigma}g_i\widehat{D}_iC^{\dagger}_{i\sigma}\widehat{Q}_{i\bar{\sigma}} = y_{i\sigma}(1-n_{i\bar{\sigma}})C^{\dagger}_{i\sigma} + y^2_{i\bar{\sigma}}y_{i\sigma}g_in_{i\bar{\sigma}}C^{\dagger}_{i\sigma} \equiv \widehat{\Gamma}_{i\sigma}C^{\dagger}_{i\sigma}$, and $g^t_{i\sigma}(T) \equiv \langle\widehat{\Gamma}_{i\sigma}\rangle^0_T/z_i(T)$. Thus we have

$$
g^t_{i\sigma}(T) = \frac{y_{i\sigma}(1-n^0_{i\bar{\sigma}}(T))}{z_i(T)} + \frac{y^2_{i\bar{\sigma}}y_{i\sigma}g_in^0_{i\bar{\sigma}}(T)}{z_i(T)} \tag{4.15}
$$

Secondly evaluate the $\cdot rhs\cdot$ of equation (4.13), that's simply $U\sum_i d_i(T) = U\sum_i \frac{(g_iy_{i\uparrow}y_{i\downarrow})^2}{z_i(T)}d^0_i(T)$.

Thus we get a renormalized non-interacting Hamiltonian $\widehat{H}_{GW}$:

$$\begin{aligned}\widehat{H}_{GW} &= -t\sum_{<ij>\sigma} g^t_{i\sigma}(T)g^t_{j\sigma}(T)c^\dagger_{i\sigma}c_{j\sigma} \\ &+U\sum_i \frac{(g_i y_{i\uparrow}y_{i\downarrow})^2 d^0_i(T)}{e^0_i(T)+y^2_{i\sigma}q^0_{i\sigma}(T)+(g_i y_{i\uparrow}y_{i\downarrow})^2 d^0_i(T)} \\ &= \widehat{H}^{renorm}_{kinetic} + U\sum_i d_i(T) \end{aligned} \tag{4.16}$$

the internal energy ($H_{int}\rho$) can now be expressed as:

$$E = (H_{int}\rho) - \mu\widehat{N} = \langle \widehat{H}_{GW}\rangle^0_T, \tag{4.17}$$

where $\langle \widehat{H}_{GW}\rangle^0_T = \frac{(\widehat{H}_{GW}e^{-\beta(\widehat{H}_0-\mu\widehat{N})})}{(e^{-\beta(\widehat{H}_0-\mu\widehat{N})})}$.

Note that $\widehat{H}_0$ can be any non-interacting Hamiltonian. We're free to take it to be $\widehat{H}_{GW}$, let's do so and then this term has clear physical meaning. With the form $E = (\widehat{H}_{GW}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})})$, it's exactly the internal energy of a non-interacting system governed by $\widehat{H}_{GW}$. Even though after making this replacement, the internal energy behaves like a non-interacting one, it doesn't mean that now the interacting system is completely equivalent to a non-interacting one, since the free energy we'll get later won't be equal to that of a non-interacting system governed by $\widehat{H}_{GW}$, namely, $F \neq F_{GW}$, where we define F_{GW} as:

$$\begin{aligned} F_{GW} &= -kT\ln Z^{GW} \\ &= -kT\ln(e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}) \\ &= -kT\ln[(e^{-\beta(\widehat{H}^{renorm}_{kinetic}-\mu\widehat{N})})\cdot e^{-\beta U\sum_i d_i}] \\ &= U\sum_i d_i - kT\sum_{\vec{k}}\ln(1+e^{-\beta[E^{GW}_{kinetic}(\vec{k})-\mu]}) \end{aligned} \tag{4.18}$$

We'll see soon that the difference comes from an extra entropy ΔS (entropy will

differ from non-interacting system's entropy by ΔS). That is to say, we use a non-interacting system governed by $\widehat{H}_{GW}$ to approximate the coherent part of the interacting system, while the effect of incoherent part is approximated by ΔS.

4.2.2 Entropy

We already know the expression of entropy is $S = -k(\rho \ln \rho)$, where $\rho = \frac{\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P}}{Z}$, thus we can further write S as:

$$S = -k\frac{[\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P}\ln(\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P})]}{Z} + k\ln Z \tag{4.19}$$

The first term in the $\cdot rhs\cdot$ of equation (4.19) equals to:

$$\begin{aligned} &-k\frac{[(\widehat{P}\ln(\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P})\widehat{P})e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}]}{Z^0\cdot\langle\widehat{P}^2\rangle_T^0} \\ &= \frac{\langle\widehat{P}\ln(\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P})\widehat{P}\rangle_T^0}{\langle\widehat{P}^2\rangle_T^0} \end{aligned} \tag{4.20}$$

Since the projection operator acts on a non-interacting eigenstate $|\Psi_n^0\rangle$ to reduce its weight if only $\langle\Psi_n^0|\widehat{D}|\Psi_n^0\rangle \neq 0$, if $\langle\Psi_n^0|\widehat{D}|\Psi_n^0\rangle = 0$, the projection keeps its weight unchanged, so it's always true that $\ln(\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P}) \leq \ln(e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}) = -\beta(\widehat{H}_{GW}-\mu\widehat{N})$ (Definitely the original Gutzwiller projector itself cannot satisfy this, which implies that normalization of the projector is required, we'll come back to discuss the normalization factor later.)

So now we have a lower bound of entropy:

$$S \geq \frac{1}{T}(\frac{\langle\widehat{P}\widehat{H}_{GW}\widehat{P}\rangle_T^0}{\langle\widehat{P}^2\rangle_T^0} - \mu\widehat{N}) + k\ln Z \tag{4.21}$$

We must be careful when evaluating $\frac{\langle\widehat{P}\widehat{H}_{GW}\widehat{P}\rangle_T^0}{\langle\widehat{P}^2\rangle_T^0}$.
Recall that $\langle\widehat{H}_{GW}\rangle_T^0 = \frac{\langle\widehat{P}(\widehat{H}_{kinetic}+U\widehat{D})\widehat{P}\rangle_T^0}{\langle\widehat{P}^2\rangle_T^0}$, one may naively think that $\frac{\langle\widehat{P}\widehat{H}_{GW}\widehat{P}\rangle_T^0}{\langle\widehat{P}^2\rangle_T^0}$ is

equal to the sum of $-t\sum_{<ij>\sigma}(g_{i\sigma}^t(T)g_{j\sigma}^t(T))^2\langle c_{i\sigma}^\dagger c_{j\sigma}\rangle_T^0$ and $U\sum_i \frac{(g_i y_{i\uparrow} y_{i\downarrow})^2 d_i^0(T)}{e_i^0(T)+y_{i\sigma}^2 q_{i\sigma}^0(T)+(g_i y_{i\uparrow} y_{i\downarrow})^2 d_i^0(T)}$. However, this is not the case. Note that $\widehat{P}$ is a projection operator. Recall the ground state Gutzwiller projection, by doing projection on the non-interacting state we are looking for a projected state $\widehat{P}|\Psi_0^0\rangle$ which can best approximate the interacting state, after we obtain it through variational process, we cannot continue to do projection on it to further reduce double occupation, otherwise we get result without physical meaning. Thus it requires us to manually add a constraint: After we specified $\widehat{P}$ and determine $\widehat{P}|\Psi_0^0\rangle$, we cannot change it further by making more projections, that is to say, $(\widehat{P})^n\widehat{P}|\Psi_0^0\rangle \quad \forall n \in \mathbb{Z}^+$, gives the same result $\widehat{P}|\Psi_0^0\rangle$. As a result, $\frac{\langle\widehat{P}\widehat{H}_{GW}\widehat{P}\rangle}{\langle\widehat{P}^2\rangle} = \langle\widehat{H}_{GW}\rangle$. Here it's the same, we have $\frac{\langle\widehat{P}\widehat{H}_{GW}\widehat{P}\rangle_T^0}{\langle\widehat{P}^2\rangle_T^0} = \langle\widehat{H}_{GW}\rangle_T^0$.

So the correct lower bound of entropy is:

$$\begin{aligned} S &\geq \frac{1}{T}(\langle\widehat{H}_{GW}\rangle_T^0 - \mu\widehat{N}) + k\ln Z \\ &= \frac{1}{T}(\langle\widehat{H}_{GW}\rangle_T^0 - \mu\widehat{N}) + k\ln(e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}) \\ &\quad + k\ln\langle\widehat{P}^2\rangle_T^0 \end{aligned} \tag{4.22}$$

In fact, if $\widehat{P}$ is a real projector satisfying $\widehat{P}^2 = \widehat{P}$, the statement above will be obvious. To be more specific, if $\widehat{P}^2 = \widehat{P}$, we can find a set of eigenstates $\{|k\rangle\}$ of $\widehat{\rho}$, it's clear that $\widehat{P}|k\rangle = |k\rangle$. Thus the numerator of the first term in the $\cdot rhs\cdot$ of

equation (4.19) can be expressed as:

$$
\begin{aligned}
&[\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P}\ln(\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P})] \\
&= \sum_k \langle k|e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}|k\rangle \ln\langle k|e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}|k\rangle \\
&\leq \sum_k \langle k|e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\ln e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}|k\rangle \\
&= (\widehat{P}e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\ln e^{-\beta(\widehat{H}_{GW}-\mu\widehat{N})}\widehat{P}) \\
&= Z^0\cdot\langle\widehat{H}_{GW}\widehat{P}^2\rangle_T^0 \\
&= Z^0\cdot\langle\widehat{H}_{GW}\rangle_T^0\cdot\langle\widehat{P}^2\rangle_T^0
\end{aligned}
\tag{4.23}
$$

Thus the first term in the $\cdot rhs\cdot$ of equation (4.19) equals to $\frac{Z^0\cdot\langle\widehat{H}_{GW}\rangle_T^0\cdot\langle\widehat{P}^2\rangle_T^0}{Z^0\cdot\langle\widehat{P}^2\rangle_T^0} = \langle\widehat{H}_{GW}\rangle_T^0$. Giving the same result.

Note that $\rho = \frac{\widehat{P}e^{-\beta(\widehat{H}_0-\mu\widehat{N})}\widehat{P}}{(\widehat{P}e^{-\beta(\widehat{H}_0-\mu\widehat{N})}\widehat{P})}$, which means if we rescale the projection operator $\widehat{P}\longrightarrow\alpha\widehat{P}$, the density matrix ρ will remain the same, so the thermal average of all observable quantities will remain the same. It seems we have full freedom to rescale $\widehat{P}$ without making any change. However, from equation 4.22 we find that if $\widehat{P}\longrightarrow\alpha\widehat{P}$, the extra entropy $\delta S\longrightarrow\delta S + k\ln\alpha^2$. This should not be allowed, definitely we cannot tune the entropy randomly, so we need some constraints on $\widehat{P}$. Recall that we used the property of projection operator before: $\widehat{P}^2=\widehat{P}$, even though the Gutzwiller type projection operator can never satisfy this condition, we can make it true at the average level: $\langle\widehat{P}^2\rangle_T^0 = \langle\widehat{P}\rangle_T^0$. Although it's an uncontrolled approximation, since we're always dealing with average, this constraint is to some extent enough, at least when U is not that large, the true ground state is not significantly different from the noninteracting one, this approximation is quite accurate. In another word, we should find an α, let $\widehat{P} = \alpha\prod_i(E_i + y_{i\sigma}Q_{i\sigma} +$

$g_i y_{i\uparrow} y_{i\downarrow} D_i)$, we have $\langle \widehat{P}^2 \rangle_T^0 = \langle \widehat{P} \rangle_T^0$. This gives:

$$\alpha = \prod_i \frac{e_i^0(T) + y_{i\sigma} q_{i\sigma}^0(T) + g_i y_{i\uparrow} y_{i\downarrow} d_i^0(T)}{e_i^0(T) + y_{i\sigma}^2 q_{i\sigma}^0(T) + g_i^2 y_{i\uparrow}^2 y_{i\downarrow}^2 d_i^0(T)} \tag{4.24}$$

Besides, we come back to the issue mentioned before that the projector $\widehat{P}$ should always reduce the wave function, now we can check that with such a normalization, at least under $d = 0$ (large positive U limit), $d = 0.25$ (noninteracting case), $d = 0.5$ (large negative U limit), the projector $\widehat{P}$ does reduce part of the wave function thus satisfying $\ln(\widehat{P} e^{-\beta(\widehat{H}_{GW} - \mu \widehat{N})} \widehat{P}) \leq \ln(e^{-\beta(\widehat{H}_{GW} - \mu \widehat{N})})$. (With a projector of Gutzwiller type, we cannot ensure that the projector is reducing wave function under all circumstance, such a normalization is the most we can do, it can at least ensure that our result will not deviate too far from the reality, namely no singularities or infinities.)

At last we get the expression of the upper bound of the interacting system's free energy:

$$\begin{aligned} F_{exact} \leq F(\rho) &= E - TS \\ &\leq F_{GW} - kT \sum_i \ln z_i(T) - kT \ln(\alpha^2) \\ \equiv F &= U \sum_i d_i - kT \sum_{\vec{k}} \ln(1 + e^{-\beta[E_{kinetic}^{GW}(\vec{k}) - \mu]}) \\ &- kT \sum_i \ln \frac{[e_i^0(T) + y_{i\sigma} q_{i\sigma}^0(T) + g_i y_{i\uparrow} y_{i\downarrow} d_i^0(T)]^2}{e_i^0(T) + y_{i\sigma}^2 q_{i\sigma}^0(T) + g_i^2 y_{i\uparrow}^2 y_{i\downarrow}^2 d_i^0(T)} \end{aligned} \tag{4.25}$$

4.3 First order phase transition in half-filled Hubbard model

Firstly note that when there's no magnetism, $d_i^0(T), e_i^0(T), q_{i\sigma}^0(T)$ are all independent of temperature, they will keep their zero-temperature value in all temperature, so do $z_i(T), g_{i\sigma}^t(T)$. Thus we can remove the temperature index. For uniform case at half filling, we can further remove the site index, considering

particle-hole symmetry, it's easy to see that:

$$d^0 = e^0 = q_\sigma^0 = 1/4 \tag{4.26}$$

$$y_{i\uparrow} = y_{i\downarrow} = \sqrt{\frac{1}{2d} - 1} \tag{4.27}$$

$$g_i y_{i\uparrow} y_{i\downarrow} = 1 \tag{4.28}$$

$$g_{i\sigma}^t = \sqrt{16d(1/2 - d)} \tag{4.29}$$

In order to compare to DMFT result, we use Bethe lattice density of state $N(\xi) = \frac{2}{\pi}\sqrt{1-\xi^2}$. Then we have:

$$\begin{aligned}\frac{F}{N} = Ud - 2kT \int d\xi N(\xi) \ln(1 + e^{-\beta \cdot 16d(1/2-d)\xi}) \\ -2kT \ln(\sqrt{d} + \sqrt{1/2 - d})\end{aligned} \tag{4.30}$$

In Fig 4.1 and Fig 4.2 two groups of plots F versus d are shown, we can see there're two ranges of T, each group represents a T value in each range. In the lower range, each T corresponds to two different Us, say $U_{c1}(T)$ and $U_{c2}(T)$. When $U \leq U_{c1}$ the free energy versus d has only one minimum; when $U_{c1} < U < U_{c2}$ another smaller minimum appears, so there're 2 minima in F-d figures; when $U \geq U_{c2}$ the larger minimum among the two disappears, only the smaller one stays, there's again only one minimum. This implies that there must be first order phase transition in the lower range of T. In the higher range of T, the two minima mix so whatever U is, there's only one minimum, we call that T separating the two ranges (where there is first order phase transition and there is not) as T_c.

We can find out T_c, and $U_{c1}(T)$, $U_{c2}(T)$ for each T below T_c. Also after doing Maxwell construction, we can find $U_c(T)$ for each T when there two minima in F-d figure and the two minima correspond to the same free energy (crossing point). The figure of phase transition is shown in Fig 4.3. In our result, $T_c = 0.19$ which

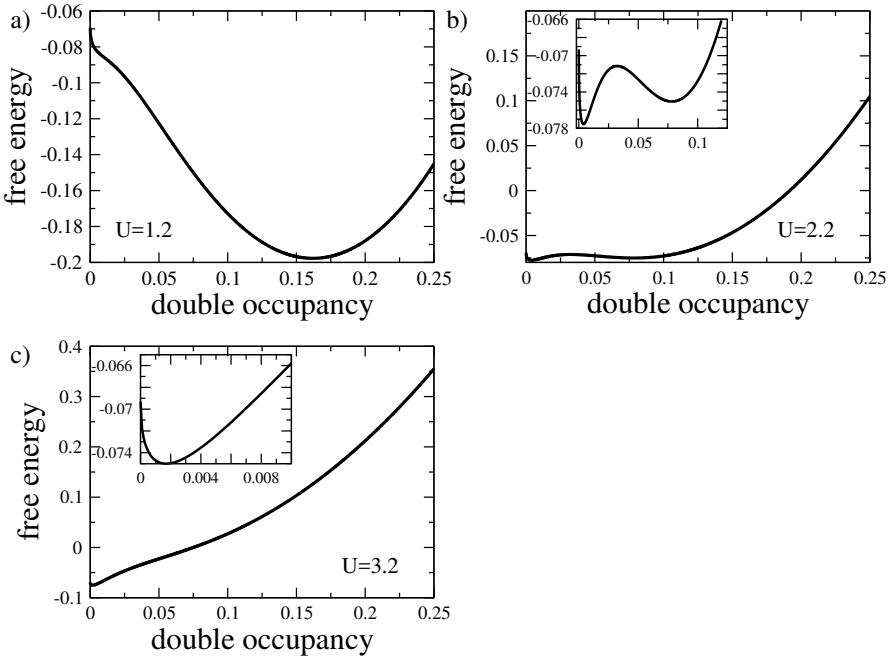

Figure 4.1: Free energy Vs double occupancy at T=0.1 for U=1.2, 2.2, 3.2. (a) $U < U_{c1}$, free energy has only one minimum at $d = 0.17$, which leads to a normal Fermi liquid. (b) $U_{c1} < U < U_{c2}$, free energy has two minima, which means two different phases coexist. (c) $U > U_{c2}$, free energy has only one mimimum at $d = 0.0015$, leading to a very narrow bandwidth which is much smaller than kT, thus giving nontrivial physical properties which will be discussed below.

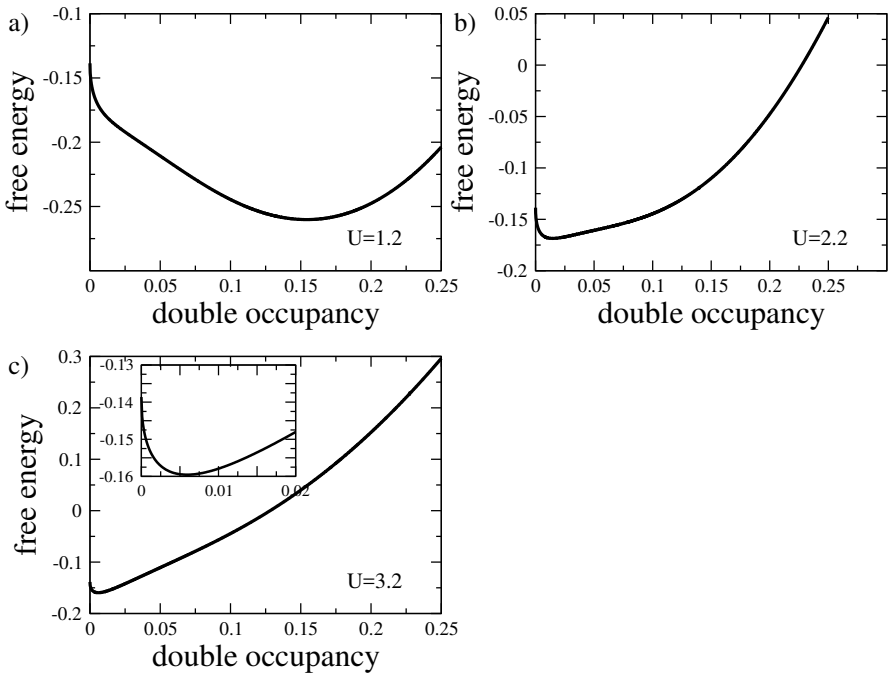

Figure 4.2: Free energy Vs double occupancy at T=0.2 for U=1.2, 2.2, 3.2. Since $T > T_c$, whatever U is, there's only one minimum in free energy, tow phases when $T < T_c$ can no longer be distinguished.

is three or four times larger than $T_c = 0.05$ in DMFT. At zero temperature, $U_c = U_{c2} = 3.4$, which is, as expected, the transition U of Brinkman-Rice transition, while U_{c1} approaches zero when temperature approaches zero.

Then we'll prove that when $T < T_c$, there's a range $(U_{c1}(T) < U < U_{c2}(T))$ where two different phases coexist (two minima),and approximately write down the equation of state of the two phases at low temperature. We start from the derivation of free energy over d.

$$\begin{aligned} F' &= \frac{\partial F}{\partial d} \\ &= U - 64(d-0.25)\int d\xi N(\xi)\frac{\xi}{e^{\frac{16d(0.5-d)\xi}{kT}}+1} \\ &\quad -kT\frac{\frac{1}{\sqrt{d}}-\frac{1}{\sqrt{0.5-d}}}{\sqrt{d}+\sqrt{0.5-d}} \end{aligned} \tag{4.31}$$

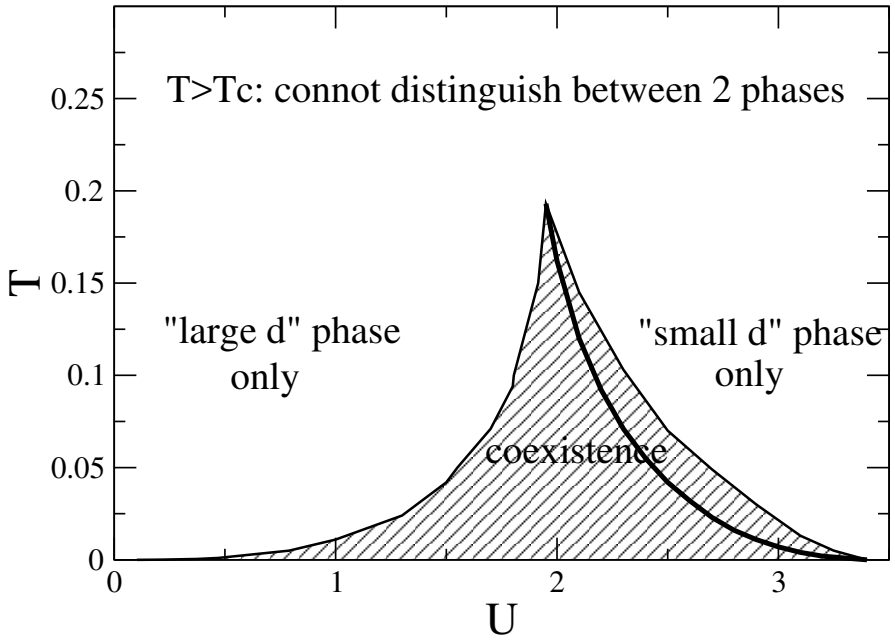

Figure 4.3: Phase diagram of half filled Hubbard model. The transition line is first order at finite temperature and terminates at 2 second order points ($U_c = 1.9, T_c = 0.19$) and ($U_c = 3.4, T_c = 0$).

It's easy to check that we can recover all $T = 0$ results. In addition, when $U = 0$ for noninteracting case, the $\cdot rhs \cdot$ of equation (4.31) is negative when $d < 0.25$, and it's zero when $d = 0.25$, and positive when $d > 0.25$, which means $d = 0.25$ is the only minimum for noninteracting case, that's what we expect.

From Fig 4.4 we can see that $F' - U$ for different temperature. When $T = 0.01$ and $T = 0.1$ (which are smaller than T_c) if $-U$ is between y axis of the two black points (whose coordinates are $(d_1(T), -U_{c1}(T))$ and $(d_2(T), -U_{c2}(T))$), then $F' = 0$ has 3 solutions for $0 \leq d \leq 0.25$, meaning that there will be 2 minima (and a maximum) in this d range. However when $T = 0.2 > T_c$, it's impossible to have 2 minima when $0 \leq d \leq 0.25$ whatever U is.

We can see that when $F' = 0$ has three solutions, one of the two solutions (except the one corresponds to maximum of free energy) is d smaller than d_1, and the other is d larger than d_2. That means, if we divide the $F' - U$ as a function of d into three parts, namely, when $d < d_1$, $d_1 \leq d \leq d_2$, and $d > d_2$, and write

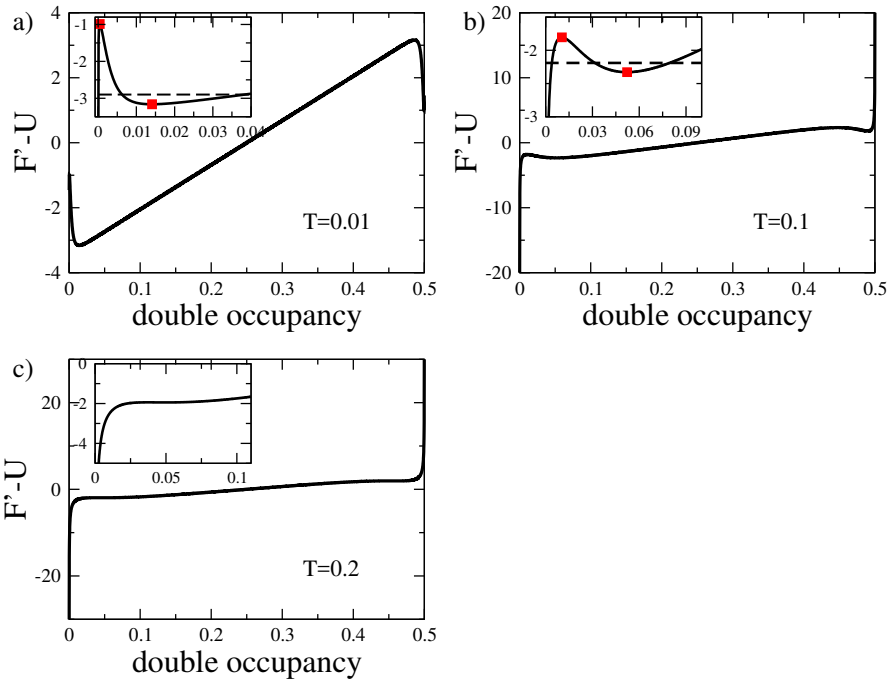

Figure 4.4: Derivative of free energy Vs double occupancy at T=0.01,0.1,0.2. In the first 2 figures, there exists a range of U which can give 3 solutions of saddle point equation $F' = 0$ when $d \leq 0.25$, while in the last figure no such U. Also in the first 2 figures we plot the Maxwell construction line, which gives the value of $-U_c(T)$.

down the approximate expression of $F' - U$ in the first and last range, we'll get equations of 'small d' state and 'large d' state. That's possible when T is small, since when T is small (e.g. $T < 0.01$), in this range, though d_1 and d_2 are both small, but we have $16d_1(0.5 - d_1)W \ll kT$ and $16d_2(0.5 - d_2)W \gg kT$ (W is bandwidth). Thus we can do high-T approximation and Sommerfeld expansion in each case.

4.3.1 $d < d_1$

In this d range, as we said, at low temperature, $16d(0.5 - d)W \ll kT$, which means we can do high-T approximation to F'.

$$\begin{aligned}
F' &= U - 64(d - 0.25)\int d\xi N(\xi)\frac{\xi}{e^{\frac{16d(0.5-d)\xi}{kT}} + 1} \\
&\quad -kT\frac{\frac{1}{\sqrt{d}} - \frac{1}{\sqrt{0.5-d}}}{\sqrt{d} + \sqrt{0.5 - d}} \\
&\approx U - 64(d - 0.25)\int d\xi N(\xi)\frac{\xi}{2(1 + \frac{8d(0.5-d)\xi}{kT})} \\
&\quad -kT\frac{\frac{1}{\sqrt{d}}}{\sqrt{0.5}} \\
&\approx U - 32(d - 0.25)\frac{2}{\pi}\int d\xi\sqrt{1 - \xi^2}\xi(1 - \frac{8d(0.5 - d)}{kT}\xi) \\
&\quad -kT\sqrt{\frac{2}{d}} \\
&= U - \frac{32d}{kT} - \sqrt{\frac{2}{d}}kT
\end{aligned} \tag{4.32}$$

So the equation of 'samll d' state is $F' = 0$, namely:

$$U = \sqrt{\frac{2}{d}}kT + \frac{32d}{kT} \tag{4.33}$$

The $\cdot rhs\cdot$ of equation (4.33) has minimum at $d = (32\sqrt{2})^{-\frac{3}{2}}(kT)^{\frac{4}{3}} = \alpha(kT)^{\frac{4}{3}}$,

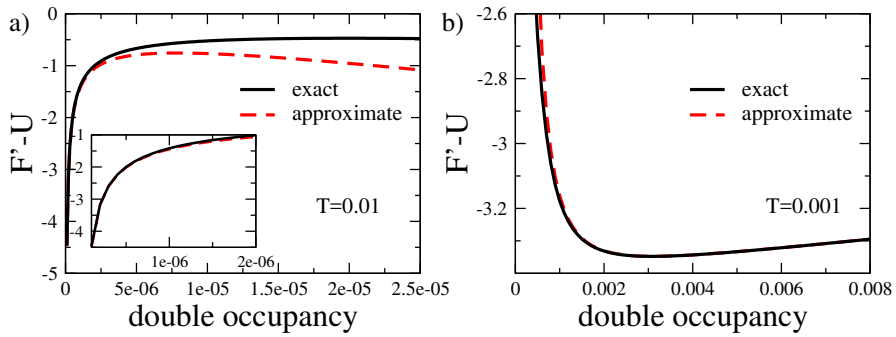

Figure 4.5: Exact and approximate forms of F' at low temperature. The left figure is when $d < d_1$ for small d range, from the inset we see the smaller d is, the better the approximate form agrees with the exact form. The right figure is when $d > d_2$ for large d range, similarly, the larger d is, the better the two forms agree with each other.

where $\alpha \approx 0.08$, the minimum value is $U_{min} = (32\alpha + \frac{\sqrt{2}}{\alpha})(kT)^{\frac{1}{3}} \approx 20.5(kT)^{\frac{1}{3}}$. Thus we get $(d_1(T), U_{c1}(T))$ at small T, which is $(\alpha(kT)^{\frac{4}{3}}, (32\alpha + \frac{\sqrt{2}}{\alpha})(kT)^{\frac{1}{3}})$. We can see that when T is small, $d_1 \propto (kT)^{\frac{4}{3}} \ll kT$, that's consistent to the requirement. Besides, the expression of U_{c1} at low temperature is:

$$U_{c1}(T) = (32\alpha + \frac{\sqrt{2}}{\alpha})(kT)^{\frac{1}{3}} \tag{4.34}$$

The left part of Fig 4.5 shows that in the $d < d_1$ range at low temperature, the approximate form of F' and the exact F' agrees very well, which justify this approximation and equation of 'small d' state.

The 'small d' implies a narrow renormalized bandwidth compared to kT. We proved analytically that for small T, the renormalized bandwidth is much smaller than kT, and in fact, according to our calculation result, even for larger T (but still smaller than T_c), the smaller value among the two d values corresponding to the two minima of free energy gives rise to a renormalized bandwidth which will be smaller than kT. It's like the high-T limit, there isn't Fermi surface anymore, here we call such phase as 'incoherent metal'.

4.3.2 $d > d_2$

In this d range, at low temperature, $16d(0.5-d)W \gg kT$, which means we can do Sommerfeld expansion to F'.

$$\begin{aligned}
F' &= U - 64(d-0.25)\int d\xi N(\xi)\frac{\xi}{e^{\frac{16d(0.5-d)\xi}{kT}}+1} \\
&\quad -kT\frac{\frac{1}{\sqrt{d}}-\frac{1}{\sqrt{0.5-d}}}{\sqrt{d}+\sqrt{0.5-d}} \\
&\approx U - 64(d-0.25)\int_{-W}^{0} d\xi N(\xi)\xi \\
&\quad -\frac{64\pi(d-0.25)}{3}[\frac{kT}{16d(0.5-d)}]^2 - \frac{\frac{1}{\sqrt{d}}-\frac{1}{\sqrt{0.5-d}}}{\sqrt{d}+\sqrt{0.5-d}}kT \\
&= U + \frac{128}{3\pi}(d-0.25) - \frac{64\pi(d-0.25)}{3}[\frac{kT}{16d(0.5-d)}]^2 \\
&\quad -\frac{\frac{1}{\sqrt{d}}-\frac{1}{\sqrt{0.5-d}}}{\sqrt{d}+\sqrt{0.5-d}}kT
\end{aligned} \tag{4.35}$$

So the equation of 'large d' state is $F' = 0$, namely:

$$\begin{aligned}
U &= -\frac{128}{3\pi}(d-0.25) + \frac{64\pi(d-0.25)}{3}[\frac{kT}{16d(0.5-d)}]^2 \\
&\quad +\frac{\frac{1}{\sqrt{d}}-\frac{1}{\sqrt{0.5-d}}}{\sqrt{d}+\sqrt{0.5-d}}kT
\end{aligned} \tag{4.36}$$

We note that even though $d > d_2$ is a large range where d can be quite large (e.g. comparable to 0.25), at low temperature d_2 itself is very small (though much larger than $\frac{kT}{W}$. Thus at the neighborhood of d_2, the derivative of free energy as a function of d can be expressed as:

$$F' = U + \frac{128}{3\pi}(d-0.25) - \sqrt{\frac{2}{d}}kT + \frac{\pi}{12}[\frac{kT}{d}]^2 \tag{4.37}$$

We see from equation (4.37) that $F'-U$ has a minimum at $d = [\frac{3(\sqrt{\frac{521}{18}}+\sqrt{\frac{1}{2}})}{\pi}]^{-\frac{2}{3}}(kT)^{\frac{2}{3}}$ $= \beta(kT)^{\frac{2}{3}}$, where $\beta \approx 0.31$. To ensure that $F' = 0$ has solution we must have $U \leq \frac{32}{3\pi} - (\frac{128\beta}{3\pi} - \sqrt{\frac{2}{\beta}} + \frac{\pi}{12\beta^2})(kT)^{\frac{2}{3}} \approx U_c(T=0) - 4.4(kT)^{\frac{2}{3}}$ (This again verifies that at zero temperature, $U_{c2} = U_c$, and when temperature is low, $U_{c2}(T)$ has a small correction $\propto (kT)^{\frac{2}{3}}$ compared to $U_{c2}(T=0)$). Thus we get $(d_2(T), U_{c2}(T))$ at small T, which is $(\beta(kT)^{\frac{2}{3}}, (\frac{32}{3\pi} - (\frac{128\beta}{3\pi} - \sqrt{\frac{2}{\beta}} + \frac{\pi}{12\beta^2})(kT)^{\frac{2}{3}})$. We can see that when T is small, $d_2 \propto (kT)^{\frac{2}{3}} \gg kT$, that's consistent to the requirement. Besides, the expression of U_{c2} at low temperature is:

$$U_{c2}(T) = \frac{32}{3\pi} - (\frac{128\beta}{3\pi} - \sqrt{\frac{2}{\beta}} + \frac{\pi}{12\beta^2})(kT)^{\frac{2}{3}} \tag{4.38}$$

The right part of Fig 4.5 shows that in the $d > d_2$ range at low temperature, the approximate form of F' and the exact F' agrees very well, which justify this approximation and equation of 'large d' state. In this state, kT is much smaller than the renormalized bandwidth, our system can still be characterized as a Fermi liquid.

Below we show the behavior of double occupancy for different U when temperature increases from infinitesimal to quite large (Ten times of half of unrenormalized bandwidth). From the figure we see if $U > U_c(T=0)$, the system always stays in the 'small d' or incoherent metal phase; while when $U_c(T_c) < U \leq U_c(T=0)$, at low temperature the 'large d' or Fermi liquid phase is favored, keeping increasing T until it touches the crossing line in the phase diagram, the first order phase transition happens, the system suddenly jumps to incoherent metal phase; and if $U < U_c(T_c)$, the Fermi liquid phase is always favored so there isn't phase transition.

At last we come to entropy. As mentioned before, the interacting system is approximated by a non-interacting system governed by H_{GW} plus effects caused by an extra entropy. From equation (4.30) we know $\Delta S = k\ln[(\sqrt{d}+\sqrt{1/2-d})^2]$.

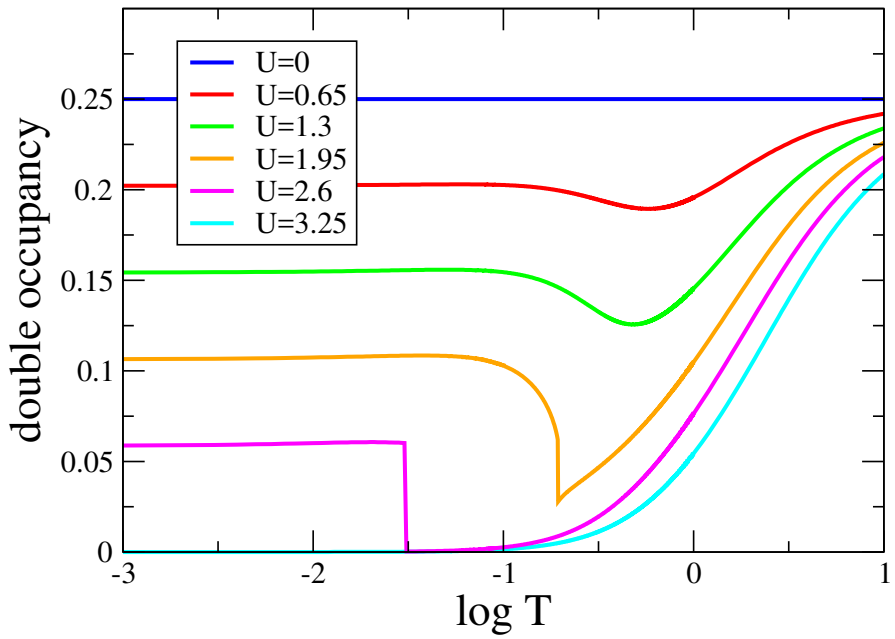

Figure 4.6: Temperature dependence of double occupancy in half filled Hubbard model, lines correspond to $U = 0, 0.65, 1.3, 1.95, 2.6, 3.25$.

So when $d = 0$, it gives an extra $-k \ln 2$. And the entropy of free Fermion part is $S_0 = 2k \sum_{\vec{k}} [\ln(1 + e^{-\beta \cdot 16d(1/2-d)E_{\vec{k}}}) + \frac{\beta E_{\vec{k}}}{e^{\beta \cdot 16d(1/2-d)E_{\vec{k}}}+1}$. When $d = 0$, $S_0 = 2k \ln 2$, which is the result of flat band. Thus $S_0 + \Delta S$ gives a total $\ln 2$ entropy, which is quite physical, since when $d = 0$ there're only 2 possible states $|\uparrow\rangle$ and $|\downarrow\rangle$ for each site.

However, when $T = 0$, $S_0 = 0$, the free Fermion part has no entropy at zero temperature. while $\Delta S = k \ln[(\sqrt{d} + \sqrt{1/2 - d})^2] \leq 0$, '=' exits only when $d = 1/4$ for non-interacting ($U = 0$) case, the total entropy will be negative. That's because the total entropy here is exactly a lower bound of the real entropy. The Figure is shown below.

4.3.3 Conclusion

To conclude, we studied metal-insulator transition in Hubbard model at half filling. We find using Gutzwiller approximation we can recover first order tran-

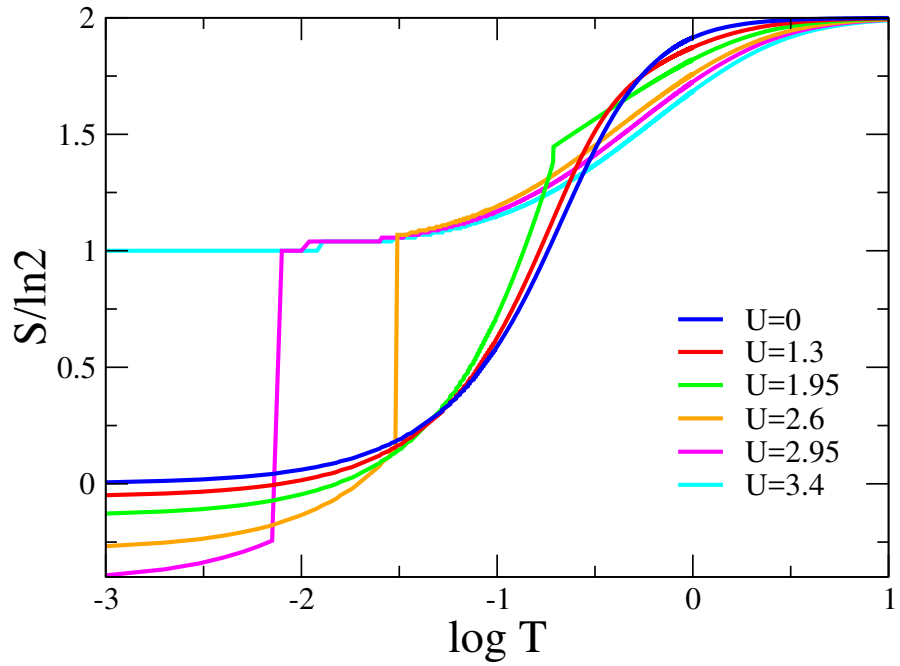

Figure 4.7: Entropy versus temperature when $U = 0, 0.65, 1.3, 1.95, 2.6, 3.4$.

sition as it is in DMFT. However the 'insulator' here does not indicate a real insulator (no double occupancy or opened gap), but a phase with very small double occupancy and thus narrow renormalized bandwidth (compared to kT), which we call 'incoherent metal'. We also get equation of state for both metal (Fermi liquid) and incoherent metal phase at low temperature, as well as the $U - T$ relationship of phase boundary at low temperature. Problems still exist, though we can qualitatively recover DMFT result, we don't get a real insulating phase, and our T_c is a bit too large compared to DMFT, and negative entropy at low temperature is a inherent problem with this method.

4.4 Antiferromagnetic phase in half filled Hubbard model

Assuming that the system is uniform, at half filling, $n_i^0(T) = 1$, so $n_{i\uparrow}^0(T) = 0.5 + m(T)$ and $n_{i\downarrow}^0(T) = 0.5 - m(T)$. For simplicity, we drop off (T) and subindex i for each physical quantity, but we keep in mind that they're self consistently

related to temperature. From equation (4.15) we can calculate Gutzwiller factor for magnetic case, specifically here we have:

$$
\begin{aligned}
g_{i\sigma}^{t} &= \frac{y_{i\sigma}(1-n_{i\bar{\sigma}}^{0}) + y_{i\bar{\sigma}}^{2} y_{i\sigma} g_i n_{i\bar{\sigma}}^{0}}{e_i^0 + y_{i\sigma}^2 q_{i\sigma}^0 + (g_i y_{i\uparrow} y_{i\downarrow})^2 d_i^0} \\
&= [\frac{(n_{i\sigma}^0 - d_i)(1 + d_i - n_i^0)}{n_{i\sigma}^0 (1 - n_{i\sigma}^0)}]^{\frac{1}{2}} + [\frac{d_i (n_{i\bar{\sigma}}^0 - d_i)}{n_{i\sigma}^0 (1 - n_{i\sigma}^0)}]^{\frac{1}{2}} \\
&= [\frac{d(0.5 + m - d)}{0.25 - m^2}]^{\frac{1}{2}} + [\frac{d(0.5 - m - d)}{0.25 - m^2}]^{\frac{1}{2}} \\
&= g_{i\bar{\sigma}}^{t} \equiv g
\end{aligned}
\tag{4.39}
$$

Dividing the whole lattice into two sublattices A and B, the Hamiltonian in antiferromagnetic case is:

$$
\begin{aligned}
\widehat{H} = -4t \sum_{\substack{\vec{k}\in RBZ \\ \sigma}} & g^2 \cos\frac{k_x}{2} \cos\frac{k_y}{2} C_{A\sigma\vec{k}}^{\dagger} C_{B\sigma\vec{k}} \\
& -\mu \sum_{\substack{\vec{k}\in RBZ \\ \sigma}} (\hat{n}_{A\sigma\vec{k}} + \hat{n}_{B\sigma\vec{k}} - 1) \\
& + \sum_{\vec{k}\in RBZ} \lambda[\hat{n}_{A\uparrow\vec{k}} + \hat{n}_{B\downarrow\vec{k}} - 2(\frac{1}{2} + m)] \\
= & \sum_{\vec{k}} (C_{A\uparrow\vec{k}}^{\dagger} \quad C_{B\uparrow\vec{k}}^{\dagger}) \quad \mathbb{M} \quad \begin{pmatrix} C_{A\uparrow\vec{k}} \\ C_{B\uparrow\vec{k}} \end{pmatrix} \\
& + \sum_{\vec{k}} (C_{B\downarrow\vec{k}}^{\dagger} \quad C_{A\downarrow\vec{k}}^{\dagger}) \quad \mathbb{M} \quad \begin{pmatrix} C_{B\downarrow\vec{k}} \\ C_{A\downarrow\vec{k}} \end{pmatrix} \\
& - \sum_{\vec{k}} [\lambda(1 + 2m) - 2\mu] \\
= & \sum_{\vec{k}\sigma} (d_{-\sigma\vec{k}}^{\dagger} \quad d_{+\sigma\vec{k}}^{\dagger}) \begin{pmatrix} E_{-\vec{k}} & 0 \\ 0 & E_{+\vec{k}} \end{pmatrix} \begin{pmatrix} d_{-\sigma} \\ d_{+\sigma} \end{pmatrix} \\
& - \sum_{\vec{k}} [\lambda(1 + 2m) - 2\mu]
\end{aligned}
\tag{4.40}
$$

where $\mathbb{M} = \begin{pmatrix} -\mu+\lambda & -4tg^2\cos\frac{k_x}{2}\cos\frac{k_y}{2} \\ -4tg^2\cos\frac{k_x}{2}\cos\frac{k_y}{2} & -\mu \end{pmatrix}$,
$E_\pm = \pm\sqrt{(8tg^2\cos\frac{k_x}{2}\cos\frac{k_y}{2})^2+\frac{\lambda^2}{4}}+\frac{\lambda}{2}-\mu$, and $\mu=\frac{\lambda}{2}$ because of particle-hole symmetry. $d_{\_\sigma\vec{k}}^{+}$ are linear combination of $C_{A\sigma\vec{k}}$ and $C_{B\sigma\vec{k}}$.

$$C_{A\uparrow} = \frac{a}{\sqrt{a^2+1}}d_{-\uparrow} + \frac{b}{\sqrt{b^2+1}}d_{+\uparrow} \tag{4.41}$$

$$C_{B\uparrow} = \frac{1}{\sqrt{a^2+1}}d_{-\uparrow} + \frac{1}{\sqrt{b^2+1}}d_{+\uparrow} \tag{4.42}$$

$$C_{B\downarrow} = \frac{a}{\sqrt{a^2+1}}d_{-\downarrow} + \frac{b}{\sqrt{b^2+1}}d_{+\downarrow} \tag{4.43}$$

$$C_{A\downarrow} = \frac{1}{\sqrt{a^2+1}}d_{-\downarrow} + \frac{1}{\sqrt{b^2+1}}d_{+\downarrow} \tag{4.44}$$

where $a = \frac{\lambda-\sqrt{(8tg^2\cos\frac{k_x}{2}\cos\frac{k_y}{2})^2+\lambda^2}}{-8tg^2\cos\frac{k_x}{2}\cos\frac{k_y}{2}}$, and $b = \frac{\lambda+\sqrt{(8tg^2\cos\frac{k_x}{2}\cos\frac{k_y}{2})^2+\lambda^2}}{-8tg^2\cos\frac{k_x}{2}\cos\frac{k_y}{2}}$.

$$\begin{aligned} m &= \frac{n_{A\uparrow}-n_{A\downarrow}}{2} = \frac{n_{B\downarrow}-n_{B\uparrow}}{2} \\ &= \frac{1}{2}[\frac{a^2-1}{a^2+1}f(E_{-\vec{k}}) + \frac{b^2-1}{b^2+1}f(E_{+\vec{k}})] \end{aligned} \tag{4.45}$$

where $f(E_{\pm\vec{k}}) = \langle d_{\pm\sigma}^\dagger d_{\pm\sigma}\rangle = \frac{1}{e^{\beta E_{\pm\vec{k}}}+1}$.

Now we get the average free energy of each electron in the projected antiferromagnetic system is:

$$\begin{aligned} f &= \frac{F}{2N} = Ud \\ &-2kT\frac{1}{2N}\sum_{\vec{k}}[\ln(1+e^{-\beta E_{-\vec{k}}}) + \ln(1+e^{-\beta E_{+\vec{k}}})] \\ &-\lambda m - kT\ln\frac{[e^0+y_\sigma q_\sigma^0 + gy_\uparrow y_\downarrow d^0]^2}{e^0+y_\sigma^2 q_\sigma^0 + (gy_\uparrow y_\downarrow)^2 d^0} \end{aligned} \tag{4.46}$$

where $d^0 = e^0 = (\frac{1}{2}+m)(\frac{1}{2}-m)$, $q_\uparrow^0 = (\frac{1}{2}+m)^2$, $q_\downarrow^0 = (\frac{1}{2}-m)^2$,

$y_\uparrow = \sqrt{\frac{(\frac{1}{2}+m-d)(\frac{1}{4}-m^2)}{(\frac{1}{2}+m)^2 d}}$, $y_\downarrow = \sqrt{\frac{(\frac{1}{2}-m-d)(\frac{1}{4}-m^2)}{(\frac{1}{2}-m)^2 d}}$, and $g y_\uparrow y_\downarrow = 1$.

4.4.1 Phase Transition

We know that within Hartree Fock mean field method, transition temperature from antiferromagnetic phase to paramagnetic phase (Neel temperature) keeps increasing while the repulsion U is increasing, which means that Hartree Fock mean field method cannot correctly characterize strong correlated system since in actual case T_{Neel} will drop when U keeps increasing. With our Gutzwiller projection method, the phase diagram and Neel temperature qualitatively agrees with other numerical calculation results such as DMFT.

However, if we look deeper into Fig 4.8 the phase diagram we'll find that when $U \leq 1.4$, if temperature increases, there will be an ordinary second order transition from AF to PM, when $1.4 < U \leq 2.0$, however, the transition turns to be a first order transition, the red line between the square dot and the circle dot is the level crossing line of the phase transition and the corresponding blue lines in the same U range describe the boundary of coexistence of the two phases. In this range T_{Neel} still increases while U increases. It's when $U > 2.0$ we see the decrease of T_{Neel}. In this range it's still first order transition, and one side of the coexistence boundary below which the PM phase does not exist is exactly the transition line between metal and incoherent metal we got previously (the green line in the figure). We'll show why it is then.

We show free energy versus magnetization at different temperature for $U \leq 1.4$, $1.4 < U \leq 2.0$ and $U > 2.0$. Although we have 3 parameters: m(magnetization), d(double occupancy) and Lagrangian multiplier λ, λ should be self-consistently determined when m and d are fixed while d is chosen to minimize free energy when m is given. Thus we can plot $F - m$ at different U and T.

In the left two plot in Fig 4.9, we can see that free energy and its derivative

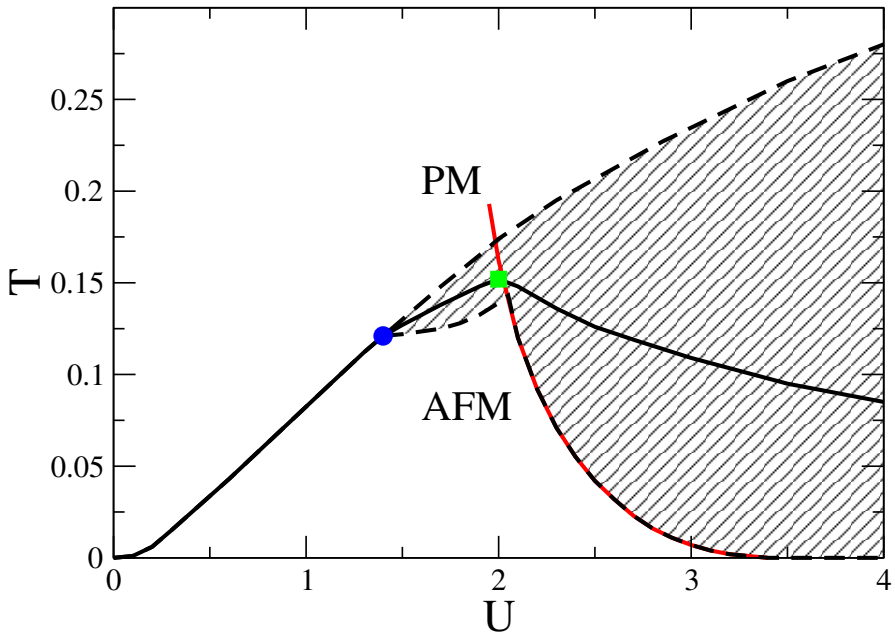

Figure 4.8: Phase diagram of AFM and PM. The thick black line is transition line between AFM and PM, left to the blue dot it's second order transtion, right to the blue dot it's first order transition. Between the blue dot and the green square, the both AFM and PM phase exist in "large d" phase, while right to the green square the first order transition is not only from AFM to PM, but from AFM "large d phase" to PM "small d incoherent metal". The shadowed area is where PM and AFM coexist, two dashed black lines are boundaries of coexistent area. The red line is transition line from metal to incoherent metal, part of which coincides with one boundary of coexistent area.

are continuous, which means when m varies, d which minimize free energy varies continuously, there is not a jump in d. We also find that d we get in these two U ranges are not too small, by saying this we mean that we stay in coherent metal phase no matter what the value of m is. The only difference between these two plots is that in the first one, m that minimize free energy changes continuously from a finite value to zero as temperature increases, there's always only one minimum in the $F - m$ plot. While in the second plot, there exists a temperature range where there're two minima, m jumps from a finite value to zero when the free energy corresponding to the two minima crosses.

In the third plot in Fig 4.9, however, when temperature is larger than a certain value, singularity appears in $F - m$ plot, which means there must be a d jump, one of the two minima belongs to metal phase while the other is bad metal.

The reason for d jump is that, if we plot free energy in $m - d$ plane, for a given m, there in general two d leading to local minimum of free energy, the local minimum corresponding to smaller d is $F - m$ in incoherent metal phase, while the other in normal metal. For $U \leq 2.0$, the $F - m$ in incoherent metal phase is always larger than that in metal for all m, that's why it's invisible to us, but when $U > 2.0$ we know from previous chapter that at least when $m = 0$, free energy for incoherent metal phase is smaller than for metal phase in appropriate temperature range. If we plot $F - m$ in both incoherent and coherent meatal phase, we'll see how they cross. The left figure in Fig 4.10 is when temperature smaller than transition temperature from metal to incoherent metal, which is $T = 0.042$ at $U = 2.5$, and the right one is when $T > 0.042$.

We examined all $F - m$ in incoherent metal phase and find that there's always only one minimum at $m = 0$, which means in incoherent metal phase there cannot exist magnetic order, in another word, incoherent metal cannot appear when the system has AFM order. Below we study the reason behind this nontrivial

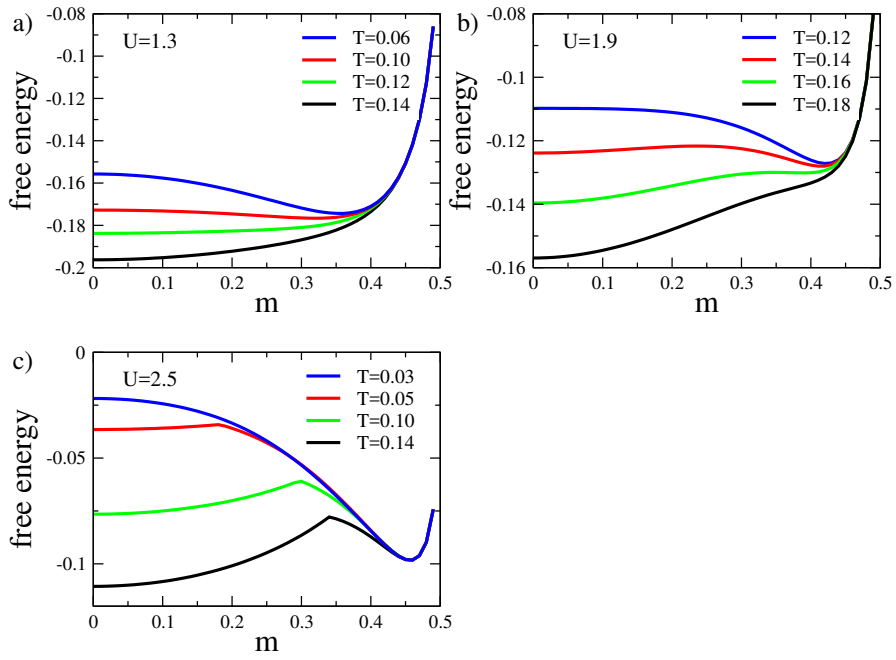

Figure 4.9: $F - m$ at different temperature for $U = 1.3$, $U = 1.9$, $U = 2.5$. In (a) we find there's always only one minimum in free energy. As temperature increases, magnetization corresponding to the minimum decreases from a finite value to zero. In (b), at low temperature there's only one minimum at finite m, but when temperature reaches a certain value, $m = 0$ gives another minimum and stays as a minimum when temperature keeps increasing and the other minimum at finite m disappears. In (c), As temperature increases, free energy becomes discontinuous when the minimum at $m = 0$ shows up, the singularity in $F - m$ implies discontinuity in double occupancy, namely, phase transition between "large d" and "small d" phases.

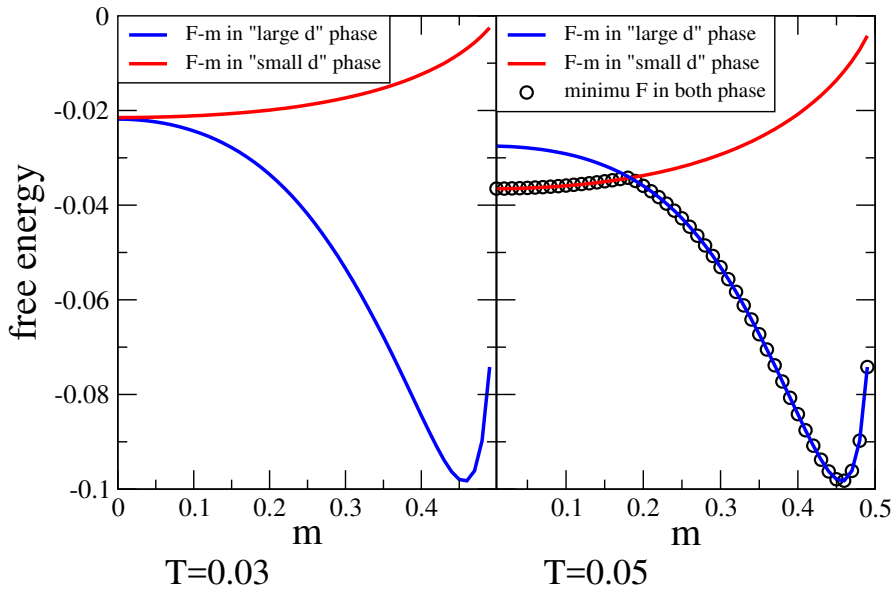

Figure 4.10: $F - m$ in both metal and incoherent metal phase at $T = 0.003$ and $T = 0.05$ when $U = 2.5$

phenomenon.

We can see from Fig 4.11 that given a small d, Gutzwiller factor is increasing with m, which means the larger m is, the smaller the kinetic part ($E_{kinetic} = -4t \sum_{\substack{\vec{k}\in RBZ \\ \sigma}} g^2 \cos\frac{k_x}{2} \cos\frac{k_y}{2} C^{\dagger}_{A\sigma\vec{k}} C_{B\sigma\vec{k}}$) of internal energy is. Meanwhile, the projective enetropy (which is always negative) is increasing with m too, thus $-TS_{projective}$ is decreasing with m, the contribution to free energy caused by projective is decreasing with m.

However the internal entropy $S_{internal}$ is decreasing with m, thus the internal entropy contribution to free energy favors small d. when$m = 0.5$, there's only one possible state thus $S_{internal} = \ln 1 = 0$, and when $m = 0$, the entropy determined by d, if d is very small (incoherent metal phase), the internal entropy is $\ln 4$ while projective entropy is $-\ln 2$, but when d is not that small (metal phase), the entropy is much smaller than $\ln 4$, as we discussed in previous chapter (See Fig 4.7. As a result, in incoherent metal phase, when m is increasing, the entropy contribution to free energy is significantly increasing, which dominates over the decrease of

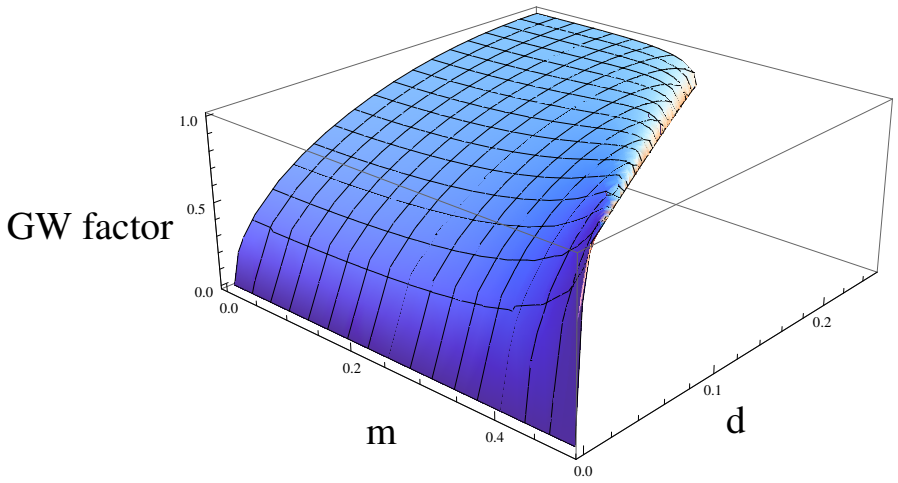

Figure 4.11: Behavior of Gutzwiller factor and projective entropy in m-d 2D plane.

kinetic energy, so paramagnetism is always favored in such case. On the other hand, in metal, the entropy contribution and kinetic contribution is comparable, they compete with each other, giving a finite m which minimizes free energy, thus magnetism shows up. We show that in Fig 4.12.

4.4.2 Conclusion

To conclude this chapter, we use projective Gutzwiller method to study magnetic phase transition. We get a Neel temperature versus repulsion U which qualitatively agrees with that of DMFT. However, with our method the drop of T_{Neel} at large U is because of a first order transition from antiferromagnetic metal to paramagnetic incoherent metal in which magnetism does not exist.

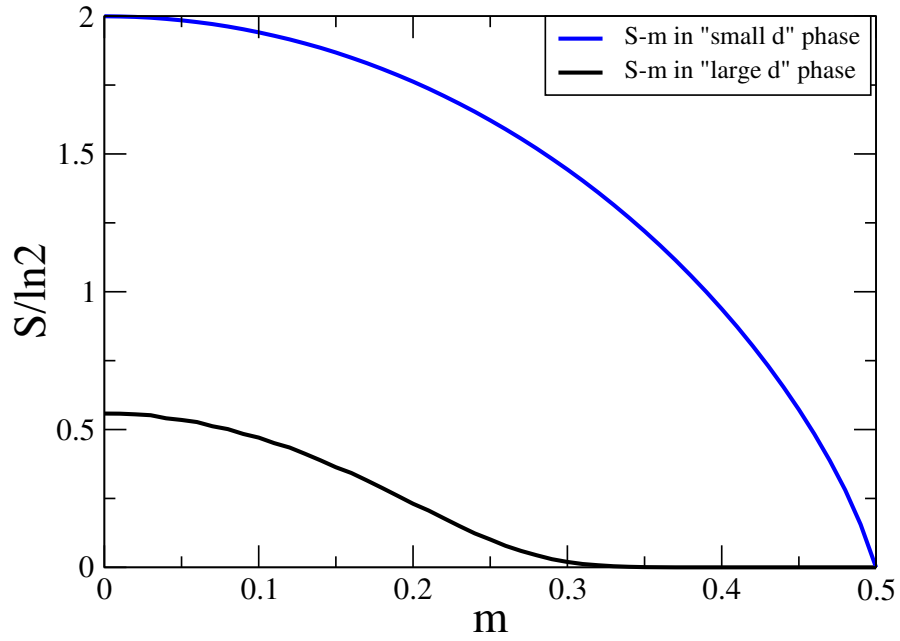

Figure 4.12: Internal entropy versus magnetization in both incoherent metal and metal phase. We can see the change of entropy in incoherent metal phase is much more drastically than that in metallic phase.

APPENDICES

APPENDIX A

Phase classification by artificial neuralnetworks

A.1 Phase classification by artificial neural networks

The artificial neural networks (NN) are computing systems widely used in data classification, pattern recognition an so on. Recently its application to condensed matter physics has been explored heavily, with significant outcomes.[49–51, 82] Here we train a one-hidden-layer NN on the same training data used in the main text, namely, the probability density of wave functions generated at small disorder strength $\delta J = 0.15 \pm 0.05$ at $\epsilon = 19/60, 59/60$, labeled as ETH, denoted by a 2 dimensional vector $(1, 0)$ and probability density of wave functions generated at large $\delta J = 9.0 \pm 1.0$ at the same energy densities, labeled as MBL, denoted by $(0, 1)$.

In the hidden layer, the inputs $\vec{x}$ are multiplied by a $2^L \times M$ dimensional matrix $\vec{W}^{(1)}$, where M is the number of nodes in the hidden layer, M ranges from 80 to 200 depending on the dimension of inputs, in another word, the size L of the spin chain. After the above linear combination, the results are added to some biases $\vec{b}^{(1)}$ and then fed to a nonlinear activation function $\Theta^{(1)}$. The work done by the

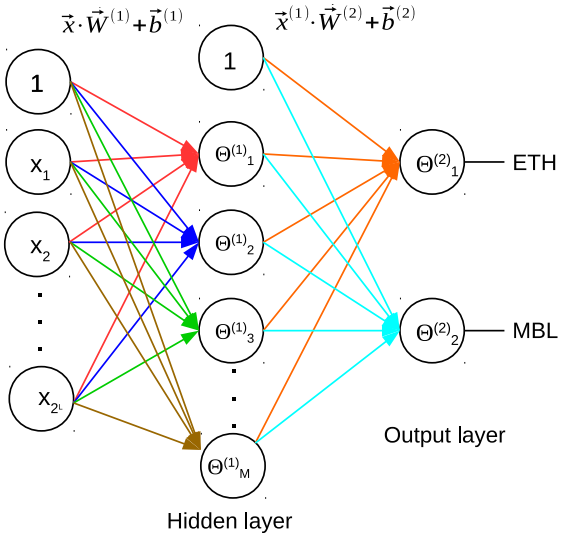

Figure A.1: Schematic explanation how NN maps an input data $\vec{x}_i$ to its label y_i, the NN acts on all input data points $i = 1, 2, \cdots, N$ thus plays a role as its target function.

first layer can be summarized as:

$$\vec{x}^{(1)} = \Theta^{(1)}(\vec{x} \cdot \vec{\vec{W}}^{(1)} + \vec{b}^{(1)}) \tag{A.1}$$

where $\Theta^{(1)}$ takes the form of ReLU[**?**], and $\vec{x}^{(1)}$ are the outputs of the hidden layer.

Similarly, the next layer, called the output layer, maps $\vec{x}^{(1)}$ to the final outputs $f(\vec{x})$ by

$$f(\vec{x}) = \Theta^{(2)}(\vec{x}^{(1)} \cdot \vec{\vec{W}}^{(2)} + \vec{b}^{(2)}) \tag{A.2}$$

where $\vec{W}^{(2)}$ is a $M \times 2$ dimensional vector performing linear combination of $\vec{x}^{(1)}$, $\vec{b}^{(2)}$ are the biases, and $\Theta^{(2)}$ takes the form of softmax[**?**] function. Thus the NN maps the 2^L dimensional inputs to 2 dimensional outputs, Fig. A.1 illustrates how NN works. The two elements of a 2 dimensional output represent the probability

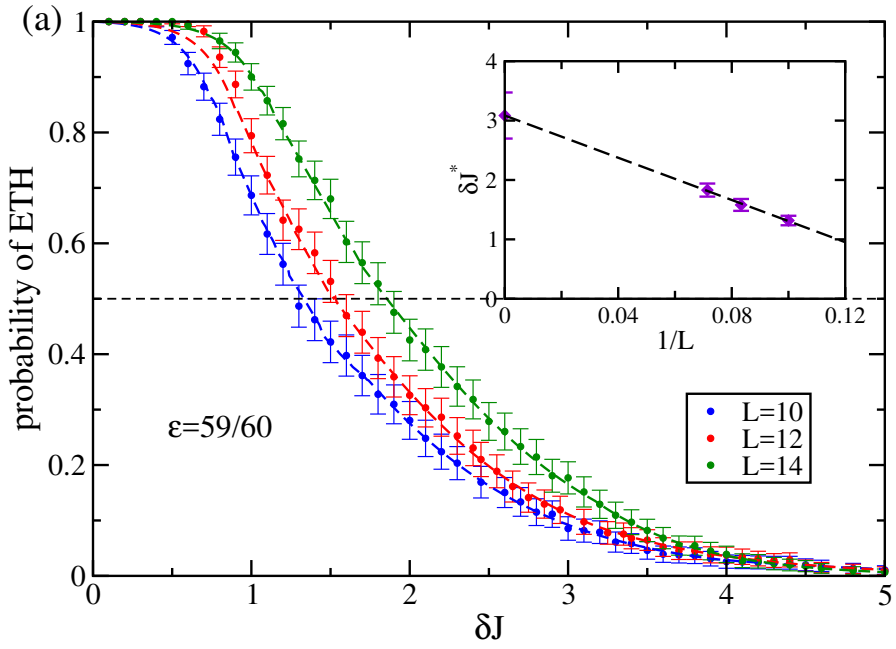

Figure A.2: The probability that eigen-wavefunction corresponding to energy density $\epsilon = 59/60$ generated at a given δJ is ETH phase for $\delta J \in [0, 5]$. The probability is estimated by the fraction of ETH phase in an ensemble of 300 disorder realizations at energy density $\epsilon = 59/60$ for $L = 10$ (blue dots), $L = 12$ (red dots) and $L = 14$ (red dots) predicted by NN. For each size, we take the δJ corresponding to 50% probability of being ETH to be the phase boundary and denote it by δJ^*. The inset shows the finite-size extrapolation of δJ^*. The intercept is interpreted as the phase boundary δJ_c in the thermodynamic limit.

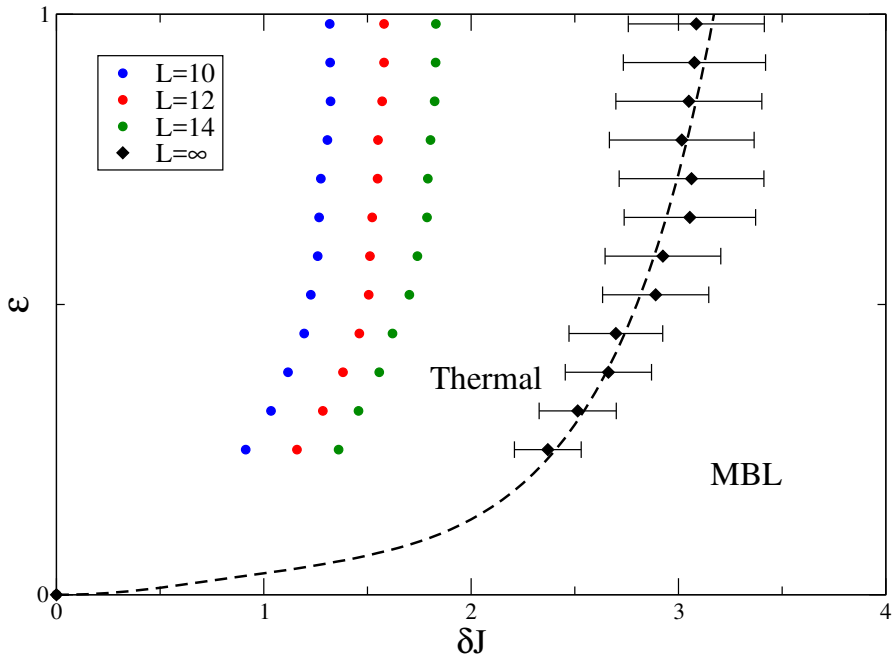

Figure A.3: Phase diagram of the disordered quantum Ising chain defined in main text. With $\epsilon = 2(E - E_{min})/(E_{max} - Emin)$ being the energy density relative to the total bandwidth. The black diamonds are δJ_c at different ϵs, which are the finite size extrapolations from the finite size transition point (blue, red and green dots).

that the input being classified as ETH and being classified as MBL respectively. The final prediction of class should be the class whose probability is larger in $f(\vec{x})$.

We use cross entropy as the cost function that acts as a metric to describe the closeness between the outputs $f(\vec{x})$ and the actual labels $\vec{y}$.

$$Cost = -\sum_{i=1}^{N} \vec{y_i} \cdot \log f(\vec{x_i}) \tag{A.3}$$

where $\vec{x_i}$ denotes input of each training sample and $\vec{y_i}$ denotes the corresponding label, N is the total number of training samples. All parameters of NN, including $\vec{\vec{W}}^{(1)}, \vec{\vec{W}}^{(2)}, \vec{b}^{(1)}, \vec{b}^{(2)}$, are determined by minimizing the cost function.

We use the same testing set as that used by SVM described in the main text. The testing accuracy is 99.8% with $L = 14$, accuracy 99.5% with $L = 12$, and

accuracy 98.8% with $L = 10$. We then follow the same procedure described in the main text to determine critical points for energy densities $\epsilon = (11 + 4i)/60, i = 1, 2, \cdots, 12$ (Fig. A.2), and then the phase boundary separating MBL and thermal phases by exponential fitting(Fig. A.3). The result obtained by using NN agrees with that of SVM within error, it also agrees with that of scaling the variance of entanglement entropy[1] within error.

Bibliography

[1] Jonas A. Kjäll, Jens H. Bardarson, and Frank Pollmann. Many-body localization in a disordered quantum ising chain. *Phys. Rev. Lett.*, 113: 107204, Sep 2014. doi: 10.1103/PhysRevLett.113.107204. URL `https://link.aps.org/doi/10.1103/PhysRevLett.113.107204`.

[2] G. Kotliar, E. Lange, and M. J. Rozenberg. Landau theory of the finite temperature mott transition. *Phys. Rev. Lett.*, 84:5180–5183, May 2000. doi: 10.1103/PhysRevLett.84.5180. URL `https://link.aps.org/doi/10.1103/PhysRevLett.84.5180`.

[3] Rahul Nandkishore and David A. Huse. Many-body localization and thermalization in quantum statistical mechanics. *Annual Review of Condensed Matter Physics*, 6(1):15–38, 2015. doi: 10.1146/annurev-conmatphys-031214-014726.

[4] Rahul Nandkishore, Sarang Gopalakrishnan, and David A. Huse. Spectral features of a many-body-localized system weakly coupled to a bath. *Phys. Rev. B*, 90:064203, Aug 2014. doi: 10.1103/PhysRevB.90.064203. URL `https://link.aps.org/doi/10.1103/PhysRevB.90.064203`.

[5] David A. Huse, Rahul Nandkishore, Vadim Oganesyan, Arijeet Pal, and S. L. Sondhi. Localization-protected quantum order. *Phys. Rev. B*, 88:014206, Jul 2013. doi: 10.1103/PhysRevB.88.014206. URL `https://link.aps.org/doi/10.1103/PhysRevB.88.014206`.

[6] D.M. Basko, I.L. Aleiner, and B.L. Altshuler. Metal–insulator transition in a weakly interacting many-electron system with localized single-particle states. *Annals of Physics*, 321(5):1126 – 1205, 2006. ISSN 0003-4916. doi: https://doi.org/10.1016/j.aop.2005.11.014. URL `http://www.sciencedirect.com/science/article/pii/S0003491605002630`.

[7] Martin C. Gutzwiller. Effect of correlation on the ferromagnetism of transition metals. *Phys. Rev. Lett.*, 10:159–162, Mar 1963. doi: 10.1103/PhysRevLett.10.159. URL `https://link.aps.org/doi/10.1103/PhysRevLett.10.159`.

[8] Martin C. Gutzwiller. Effect of correlation on the ferromagnetism of transition metals. *Phys. Rev.*, 134:A923–A941, May 1964. doi: 10.1103/PhysRev.134.A923. URL `https://link.aps.org/doi/10.1103/PhysRev.134.A923`.

[9] Martin C. Gutzwiller. Correlation of electrons in a narrow s band. *Phys. Rev.*, 137:A1726–A1735, Mar 1965. doi: 10.1103/PhysRev.137.A1726. URL `https://link.aps.org/doi/10.1103/PhysRev.137.A1726`.

[10] A. Mildenberger, F. Evers, and A. D. Mirlin. Dimensionality dependence of the wave-function statistics at the anderson transition. *Phys. Rev. B*, 66:033109, Jul 2002. doi: 10.1103/PhysRevB.66.033109. URL `https://link.aps.org/doi/10.1103/PhysRevB.66.033109`.

[11] V. Dobrosavljević, A. A. Pastor, and B. K. Nikolić. Typical medium theory of Anderson localization: A local order parameter approach to strong-disorder effects. *Europhys. Lett.*, 62:76, 2003.

[12] Hanna Terletska, Yi Zhang, Ka-Ming Tam, Tom Berlijn, Liviu Chioncel, N. S. Vidhyadhiraja, and Mark Jarrell. Systematic quantum cluster typical medium method for the study of localization in strongly disordered electronic systems.

Applied Sciences, 8(12), 2018. ISSN 2076-3417. doi: 10.3390/app8122401. URL http://www.mdpi.com/2076-3417/8/12/2401.

[13] Yi Zhang, R. Nelson, Elisha Siddiqui, K.-M. Tam, U. Yu, T. Berlijn, W. Ku, N. S. Vidhyadhiraja, J. Moreno, and M. Jarrell. Generalized multiband typical medium dynamical cluster approximation: Application to (ga,mn)n. *Phys. Rev. B*, 94:224208, Dec 2016. doi: 10.1103/PhysRevB.94.224208. URL https://link.aps.org/doi/10.1103/PhysRevB.94.224208.

[14] C. E. Ekuma, V. Dobrosavljević, and D. Gunlycke. First-principles-based method for electron localization: Application to monolayer hexagonal boron nitride. *Phys. Rev. Lett.*, 118:106404, Mar 2017. doi: 10.1103/PhysRevLett.118.106404. URL https://link.aps.org/doi/10.1103/PhysRevLett.118.106404.

[15] Yi Zhang, R. Nelson, K.-M. Tam, W. Ku, U. Yu, N. S. Vidhyadhiraja, H. Terletska, J. Moreno, M. Jarrell, and T. Berlijn. Origin of localization in ti-doped si. *Phys. Rev. B*, 98:174204, Nov 2018. doi: 10.1103/PhysRevB.98.174204. URL https://link.aps.org/doi/10.1103/PhysRevB.98.174204.

[16] David J. Luitz, Nicolas Laflorencie, and Fabien Alet. Many-body localization edge in the random-field heisenberg chain. *Phys. Rev. B*, 91:081103, Feb 2015. doi: 10.1103/PhysRevB.91.081103. URL https://link.aps.org/doi/10.1103/PhysRevB.91.081103.

[17] Vedika Khemani, S. P. Lim, D. N. Sheng, and David A. Huse. Critical properties of the many-body localization transition. *Phys. Rev. X*, 7:021013, Apr 2017. doi: 10.1103/PhysRevX.7.021013. URL https://link.aps.org/doi/10.1103/PhysRevX.7.021013.

[18] Hubbard.J. Electron correlations in narrow energy bands. *Proceedings of the*

Royal Society of London. Series A. Mathematical and Physical Sciences, 276: 238, 1963. doi: http://doi.org/10.1098/rspa.1963.0204.

[19] Walter Metzner and Dieter Vollhardt. Ground-state properties of correlated fermions: Exact analytic results for the gutzwiller wave function. *Phys. Rev. Lett.*, 59:121–124, Jul 1987. doi: 10.1103/PhysRevLett.59.121. URL `https://link.aps.org/doi/10.1103/PhysRevLett.59.121`.

[20] F. Gebhard and D. Vollhardt. Correlation functions for hubbard-type models: The exact results for the gutzwiller wave function in one dimension. *Phys. Rev. Lett.*, 59:1472–1475, Sep 1987. doi: 10.1103/PhysRevLett.59.1472. URL `https://link.aps.org/doi/10.1103/PhysRevLett.59.1472`.

[21] J. M. Deutsch. Quantum statistical mechanics in a closed system. *Phys. Rev. A*, 43:2046–2049, Feb 1991. doi: 10.1103/PhysRevA.43.2046. URL `https://link.aps.org/doi/10.1103/PhysRevA.43.2046`.

[22] Mark Srednicki. Chaos and quantum thermalization. *Phys. Rev. E*, 50:888–901, Aug 1994. doi: 10.1103/PhysRevE.50.888. URL `https://link.aps.org/doi/10.1103/PhysRevE.50.888`.

[23] Marcos Rigol, Vanja Dunjko, and Maxim Olshanii. Thermalization and its mechanism for generic isolated quantum systems. *Nature*, 452:854 EP –, Apr 2008. URL `http://dx.doi.org/10.1038/nature06838`.

[24] Bela Bauer and Chetan Nayak. Area laws in a many-body localized state and its implications for topological order. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(09):P09005, 2013. URL `http://stacks.iop.org/1742-5468/2013/i=09/a=P09005`.

[25] Marko Žnidarič, Toma ž Prosen, and Peter Prelovšek. Many-body localization in the heisenberg xxz magnet in a random field. *Phys. Rev. B*, 77:064426,

Feb 2008. doi: 10.1103/PhysRevB.77.064426. URL https://link.aps.org/doi/10.1103/PhysRevB.77.064426.

[26] Jens H. Bardarson, Frank Pollmann, and Joel E. Moore. Unbounded growth of entanglement in models of many-body localization. *Phys. Rev. Lett.*, 109: 017202, Jul 2012. doi: 10.1103/PhysRevLett.109.017202. URL https://link.aps.org/doi/10.1103/PhysRevLett.109.017202.

[27] Ronen Vosk and Ehud Altman. Many-body localization in one dimension as a dynamical renormalization group fixed point. *Phys. Rev. Lett.*, 110:067204, Feb 2013. doi: 10.1103/PhysRevLett.110.067204. URL https://link.aps.org/doi/10.1103/PhysRevLett.110.067204.

[28] Maksym Serbyn, Z. Papić, and Dmitry A. Abanin. Universal slow growth of entanglement in interacting strongly disordered systems. *Phys. Rev. Lett.*, 110:260601, Jun 2013. doi: 10.1103/PhysRevLett.110.260601. URL https://link.aps.org/doi/10.1103/PhysRevLett.110.260601.

[29] Maksym Serbyn, Z. Papić, and Dmitry A. Abanin. Local conservation laws and the structure of the many-body localized states. *Phys. Rev. Lett.*, 111: 127201, Sep 2013. doi: 10.1103/PhysRevLett.111.127201. URL https://link.aps.org/doi/10.1103/PhysRevLett.111.127201.

[30] T. Grover. Certain General Constraints on the Many-Body Localization Transition. *ArXiv e-prints*, May 2014.

[31] J. Eisert, M. Cramer, and M. B. Plenio. Colloquium. *Rev. Mod. Phys.*, 82: 277–306, Feb 2010. doi: 10.1103/RevModPhys.82.277. URL https://link.aps.org/doi/10.1103/RevModPhys.82.277.

[32] D.M. Basko, I.L. Aleiner, and B.L. Altshuler. Metal–insulator transition in a weakly interacting many-electron system with localized single-particle states.

Annals of Physics, 321(5):1126 – 1205, 2006. ISSN 0003-4916. doi: https://doi.org/10.1016/j.aop.2005.11.014. URL http://www.sciencedirect.com/science/article/pii/S0003491605002630.

[33] Scott D. Geraedts, Rahul Nandkishore, and Nicolas Regnault. Many-body localization and thermalization: Insights from the entanglement spectrum. *Phys. Rev. B*, 93:174202, May 2016. doi: 10.1103/PhysRevB.93.174202. URL https://link.aps.org/doi/10.1103/PhysRevB.93.174202.

[34] B. Swingle. A simple model of many-body localization. *ArXiv e-prints*, July 2013.

[35] V. Ros, M. Müller, and A. Scardicchio. Integrals of motion in the many-body localized phase. *Nuclear Physics B*, 891:420 – 465, 2015. ISSN 0550-3213. doi: https://doi.org/10.1016/j.nuclphysb.2014.12.014. URL http://www.sciencedirect.com/science/article/pii/S0550321314003836.

[36] Ian Mondragon-Shem, Arijeet Pal, Taylor L. Hughes, and Chris R. Laumann. Many-body mobility edge due to symmetry-constrained dynamics and strong interactions. *Phys. Rev. B*, 92:064203, Aug 2015. doi: 10.1103/PhysRevB.92.064203. URL https://link.aps.org/doi/10.1103/PhysRevB.92.064203.

[37] Ranjan Modak and Subroto Mukerjee. Many-body localization in the presence of a single-particle mobility edge. *Phys. Rev. Lett.*, 115:230401, Dec 2015. doi: 10.1103/PhysRevLett.115.230401. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.230401.

[38] Elliott Baygan, S. P. Lim, and D. N. Sheng. Many-body localization and mobility edge in a disordered spin-$\frac{1}{2}$ heisenberg ladder. *Phys. Rev. B*, 92:195153, Nov 2015. doi: 10.1103/PhysRevB.92.195153. URL https://link.aps.org/doi/10.1103/PhysRevB.92.195153.

[39] David J. Luitz. Long tail distributions near the many-body localization transition. *Phys. Rev. B*, 93:134201, Apr 2016. doi: 10.1103/PhysRevB.93.134201. URL https://link.aps.org/doi/10.1103/PhysRevB.93.134201.

[40] Sabyasachi Nag and Arti Garg. Many-body mobility edges in a one-dimensional system of interacting fermions. *Phys. Rev. B*, 96:060203, Aug 2017. doi: 10.1103/PhysRevB.96.060203. URL https://link.aps.org/doi/10.1103/PhysRevB.96.060203.

[41] P. W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109:1492–1505, Mar 1958. doi: 10.1103/PhysRev.109.1492. URL https://link.aps.org/doi/10.1103/PhysRev.109.1492.

[42] J. Brndiar and P. Markoš. Universality of the metal-insulator transition in three-dimensional disordered systems. *Phys. Rev. B*, 74:153103, Oct 2006. doi: 10.1103/PhysRevB.74.153103. URL https://link.aps.org/doi/10.1103/PhysRevB.74.153103.

[43] Torres-Herrera E. J. and Santos Lea F. Extended nonergodic states in disordered many-body quantum systems. *Annalen der Physik*, 529(7):1600284. doi: 10.1002/andp.201600284. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.201600284.

[44] Soumya Bera, Henning Schomerus, Fabian Heidrich-Meisner, and Jens H. Bardarson. Many-body localization characterized from a one-particle perspective. *Phys. Rev. Lett.*, 115:046603, Jul 2015. doi: 10.1103/PhysRevLett.115.046603. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.046603.

[45] W Beugeling, A Andreanov, and Masudul Haque. Global characteristics of all eigenstates of local many-body hamiltonians: participation ratio and entan-

glement entropy. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(2):P02002, 2015. URL http://stacks.iop.org/1742-5468/2015/i=2/a=P02002.

[46] Xiao Chen, Benjamin Hsu, Taylor L. Hughes, and Eduardo Fradkin. Rényi entropy and the multifractal spectra of systems near the localization transition. *Phys. Rev. B*, 86:134201, Oct 2012. doi: 10.1103/PhysRevB.86.134201. URL https://link.aps.org/doi/10.1103/PhysRevB.86.134201.

[47] Giulio Biroli, Guilhem Semerjian, and Marco Tarzia. Anderson model on bethe lattices: Density of states, localization properties and isolated eigenvalue. *Progress of Theoretical Physics Supplement*, 184:187–199, 2010. doi: 10.1143/PTPS.184.187. URL http://dx.doi.org/10.1143/PTPS.184.187.

[48] P. Roushan, C. Neill, J. Tangpanitanon, V. M. Bastidas, A. Megrant, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, B. Foxen, M. Giustina, E. Jeffrey, J. Kelly, E. Lucero, J. Mutus, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. White, H. Neven, D. G. Angelakis, and J. Martinis. Spectroscopic signatures of localization with interacting photons in superconducting qubits. *Science*, 358(6367): 1175–1179, 2017. ISSN 0036-8075. doi: 10.1126/science.aao1401. URL https://science.sciencemag.org/content/358/6367/1175.

[49] Frank Schindler, Nicolas Regnault, and Titus Neupert. Probing many-body localization with neural networks. *Phys. Rev. B*, 95:245134, Jun 2017. doi: 10.1103/PhysRevB.95.245134. URL https://link.aps.org/doi/10.1103/PhysRevB.95.245134.

[50] EvertÂ?P Â?L van Nieuwenburg, Ye-Hua Liu, and SebastianÂ?D Huber.

Learning phase transitions by confusion. *Nature Physics*, 13:435 EP –, Feb 2017. URL http://dx.doi.org/10.1038/nphys4037.

[51] Juan Carrasquilla and Roger G. Melko. Machine learning phases of matter. *Nature Physics*, 13:431 EP –, Feb 2017. URL http://dx.doi.org/10.1038/nphys4035.

[52] Yi Zhang, Roger G. Melko, and Eun-Ah Kim. Machine learning $\mathbb{z}_2$ quantum spin liquids with quasiparticle statistics. *Phys. Rev. B*, 96:245119, Dec 2017. doi: 10.1103/PhysRevB.96.245119. URL https://link.aps.org/doi/10.1103/PhysRevB.96.245119.

[53] Y.-T. Hsu, X. Li, D.-L. Deng, and S. Das Sarma. Machine learning many-body localization: Search for the elusive nonergodic metal. *ArXiv e-prints*, May 2018.

[54] Lei Wang. Discovering phase transitions with unsupervised learning. *Phys. Rev. B*, 94:195105, Nov 2016. doi: 10.1103/PhysRevB.94.195105. URL https://link.aps.org/doi/10.1103/PhysRevB.94.195105.

[55] Pedro Ponte and Roger G. Melko. Kernel methods for interpretable machine learning of order parameters. *Phys. Rev. B*, 96:205146, Nov 2017. doi: 10.1103/PhysRevB.96.205146. URL https://link.aps.org/doi/10.1103/PhysRevB.96.205146.

[56] J. Greitemann, K. Liu, and L. Pollet. Probing Hidden Spin Order with Interpretable Machine Learning. *ArXiv e-prints*, April 2018.

[57] F. Wegner. Inverse participation ratio in $2+\epsilon$ dimensions. *Zeitschrift für Physik B Condensed Matter*, 36(3):209–214, Sep 1980. URL https://doi.org/10.1007/BF01325284.

[58] David J. Luitz, Fabien Alet, and Nicolas Laflorencie. Universal behavior beyond multifractality in quantum many-body systems. *Phys. Rev. Lett.*, 112:057203, Feb 2014. doi: 10.1103/PhysRevLett.112.057203. URL https://link.aps.org/doi/10.1103/PhysRevLett.112.057203.

[59] Arijeet Pal and David A. Huse. Many-body localization phase transition. *Phys. Rev. B*, 82:174411, Nov 2010. doi: 10.1103/PhysRevB.82.174411. URL https://link.aps.org/doi/10.1103/PhysRevB.82.174411.

[60] Vadim Oganesyan and David A. Huse. Localization of interacting fermions at high temperature. *Phys. Rev. B*, 75:155111, Apr 2007. doi: 10.1103/PhysRevB.75.155111. URL https://link.aps.org/doi/10.1103/PhysRevB.75.155111.

[61] Ehud Altman and Ronen Vosk. Universal dynamics and renormalization in many-body-localized systems. *Annual Review of Condensed Matter Physics*, 6(1):383–409, 2015. doi: 10.1146/annurev-conmatphys-031214-014701. URL https://doi.org/10.1146/annurev-conmatphys-031214-014701.

[62] Michael Schreiber, Sean S. Hodgman, Pranjal Bordia, Henrik P. Lüschen, Mark H. Fischer, Ronen Vosk, Ehud Altman, Ulrich Schneider, and Immanuel Bloch. Observation of many-body localization of interacting fermions in a quasirandom optical lattice. *Science*, 349(6250):842–845, 2015. ISSN 0036-8075. doi: 10.1126/science.aaa7432. URL https://science.sciencemag.org/content/349/6250/842.

[63] P. Jurcevic, H. Shen, P. Hauke, C. Maier, T. Brydges, C. Hempel, B. P. Lanyon, M. Heyl, R. Blatt, and C. F. Roos. Direct observation of dynamical quantum phase transitions in an interacting many-body system. *Phys. Rev.*

Lett., 119:080501, Aug 2017. doi: 10.1103/PhysRevLett.119.080501. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.080501.

[64] J. Smith, A. Lee, P. Richerme, B. Neyenhuis, P. W. Hess, P. Hauke, M. Heyl, D. A. Huse, and C. Monroe. Many-body localization in a quantum simulator with programmable random disorder. *Nature Physics*, 12:907 EP –, Jun 2016. URL https://doi.org/10.1038/nphys3783.

[65] Henrik P. Lüschen, Pranjal Bordia, Sean S. Hodgman, Michael Schreiber, Saubhik Sarkar, Andrew J. Daley, Mark H. Fischer, Ehud Altman, Immanuel Bloch, and Ulrich Schneider. Signatures of many-body localization in a controlled open quantum system. *Phys. Rev. X*, 7:011034, Mar 2017. doi: 10.1103/PhysRevX.7.011034. URL https://link.aps.org/doi/10.1103/PhysRevX.7.011034.

[66] Pranjal Bordia, Henrik P. Lüschen, Sean S. Hodgman, Michael Schreiber, Immanuel Bloch, and Ulrich Schneider. Coupling identical one-dimensional many-body localized systems. *Phys. Rev. Lett.*, 116:140401, Apr 2016. doi: 10.1103/PhysRevLett.116.140401. URL https://link.aps.org/doi/10.1103/PhysRevLett.116.140401.

[67] Tarun Grover. Certain General Constraints on the Many-Body Localization Transition. *arXiv e-prints*, art. arXiv:1405.1471, May 2014.

[68] Wei Zhang and Ziqiang Wang. Critical behaviors of the entanglement and participation entropy near the many-body localization transition in a disordered quantum spin chain. *arXiv e-prints*, art. arXiv:1908.04039, Aug 2019.

[69] A. Chandran, C. R. Laumann, and V. Oganesyan. Finite size scaling bounds on many-body localized phase transitions. *arXiv e-prints*, art. arXiv:1509.04285, Sep 2015.

[70] Ronen Vosk, David A. Huse, and Ehud Altman. Theory of the many-body localization transition in one-dimensional systems. *Phys. Rev. X*, 5:031032, Sep 2015. doi: 10.1103/PhysRevX.5.031032. URL https://link.aps.org/doi/10.1103/PhysRevX.5.031032.

[71] Andrew C. Potter, Romain Vasseur, and S. A. Parameswaran. Universal properties of many-body delocalization transitions. *Phys. Rev. X*, 5:031033, Sep 2015. doi: 10.1103/PhysRevX.5.031033. URL https://link.aps.org/doi/10.1103/PhysRevX.5.031033.

[72] Liangsheng Zhang, Bo Zhao, Trithep Devakul, and David A. Huse. Many-body localization phase transition: A simplified strong-randomness approximate renormalization group. *Phys. Rev. B*, 93:224201, Jun 2016. doi: 10.1103/PhysRevB.93.224201. URL https://link.aps.org/doi/10.1103/PhysRevB.93.224201.

[73] Cécile Monthus. Many-body-localization transition in the strong disorder limit: Entanglement entropy from the statistics of rare extensive resonances. *Entropy*, 18(4), 2016. ISSN 1099-4300. doi: 10.3390/e18040122. URL https://www.mdpi.com/1099-4300/18/4/122.

[74] A. A. Aligia, C. D. Batista, and F. H. L. Eßler. Numerical method for detecting incommensurate correlations in the heisenberg zigzag ladder, Aug 2000. URL https://link.aps.org/doi/10.1103/PhysRevB.62.3259.

[75] Mischa Thesberg and Erik S. Sørensen. Exact diagonalization study of the anisotropic triangular lattice heisenberg model using twisted boundary conditions. *Phys. Rev. B*, 90:115117, Sep 2014. doi: 10.1103/PhysRevB.90.115117. URL https://link.aps.org/doi/10.1103/PhysRevB.90.115117.

[76] Philipp T. Dumitrescu, Romain Vasseur, and Andrew C. Potter. Scaling

theory of entanglement at the many-body localization transition. *Phys. Rev. Lett.*, 119:110604, Sep 2017. doi: 10.1103/PhysRevLett.119.110604. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.110604.

[77] Trithep Devakul and Rajiv R. P. Singh. Early breakdown of area-law entanglement at the many-body delocalization transition. *Phys. Rev. Lett.*, 115:187201, Oct 2015. doi: 10.1103/PhysRevLett.115.187201. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.187201.

[78] S. P. Lim and D. N. Sheng. Many-body localization and transition by density matrix renormalization group and exact diagonalization studies. *Phys. Rev. B*, 94:045111, Jul 2016. doi: 10.1103/PhysRevB.94.045111. URL https://link.aps.org/doi/10.1103/PhysRevB.94.045111.

[79] Torres-Herrera E. J. and Santos Lea F. Extended nonergodic states in disordered many-body quantum systems. *Annalen der Physik*, 529(7):1600284, 2017.

[80] Wei Zhang, Lei Wang, and Ziqiang Wang. Interpretable machine learning study of the many-body localization transition in disordered quantum ising spin chains. *Phys. Rev. B*, 99:054208, Feb 2019. doi: 10.1103/PhysRevB.99.054208. URL https://link.aps.org/doi/10.1103/PhysRevB.99.054208.

[81] R J Bell. The dynamics of disordered lattices. *Reports on Progress in Physics*, 35(3):1315–1409, sep 1972. doi: 10.1088/0034-4885/35/3/306. URL https://doi.org/10.1088%2F0034-4885%2F35%2F3%2F306.

[82] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017. ISSN 0036-8075. doi: 10.1126/science.aag2302. URL http://science.sciencemag.org/content/355/6325/602.